

## AUTOMATIC TRANSLATION AND RETRANSLATION SYSTEM

**Publication number: JP2002512717T**

**Publication date:** 2002-04-23

**Inventor:**

**Applicant:**

**Classification:**


**- International:** G06F17/28; G06F17/28; (IPC1-7): G06F17/28

**- european:** G06F17/27A2; G06F17/27R4; G06F17/28R;  
G06F17/28U

**Application number: JP19990502368T.19980114**

**Priority number(s):** WO1997US10005, 19970609; WO1998US00729, 19980114

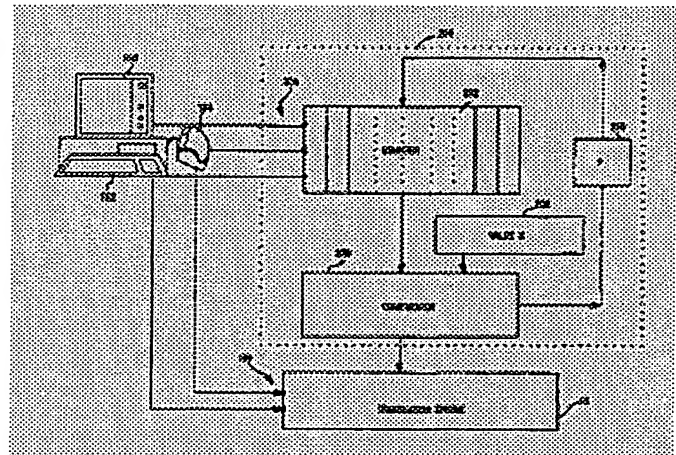
**Also published as:**

 WO9857271. (A1)

**Report a data error here**

Abstract not available for JP2002512717T.  
Abstract of corresponding document: WO9857271.

An automatic natural translation system for translating input text in a source natural language, such as English, to output text in a target natural language, such as Japanese, as the input text is being generated, includes a translation activator that determines when a pause occurs in the creation of source language input text in a document or file, and effects, in response to the pause, a translation of the input text in a target language, up to that point that the pause was sensed. The translation system can further effect a translation as text is being generated, in response to sensing a certain text structure or in response to an input function. Each time an automatic translation is performed, the translation can begin at a starting point, such as the beginning of a word processing document into which the text was entered, the beginning of immediately received input text, or the beginning of a page or paragraph, thus allowing any changes in sentence structure or recent edits to the input text to be reflected in the output text translation.



Data supplied from the **esp@cenet** database. - Worldwide.

(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2002-512717

(P2002-512717A)

(43) 公表日 平成14年4月23日 (2002.4.23)

(51) Int.Cl.<sup>7</sup>

識別記号

F I

テーマコード\* (参考)

G 0 6 F 17/28

G 0 6 F 17/28

B

審査請求 有 予備審査請求 有 (全 75 頁)

(21) 出願番号 特願平11-502368  
 (86) (22) 出願日 平成10年1月14日 (1998.1.14)  
 (85) 翻訳文提出日 平成11年12月9日 (1999.12.9)  
 (86) 国際出願番号 PCT/US 98/00729  
 (87) 国際公開番号 WO 98/57271  
 (87) 国際公開日 平成10年12月17日 (1998.12.17)  
 (31) 優先権主張番号 PCT/US 97/10005  
 (32) 優先日 平成9年6月9日 (1997.6.9)  
 (33) 優先権主張国 世界知的所有権機関 (WO)  
 (81) 指定国 JP, US

(71) 出願人 ログヴィスタ株式会社  
 東京都江東区潮見2丁目10-24  
 (72) 発明者 ブリングル, ルイス ジー.  
 アメリカ合衆国 マサチューセッツ  
 01776, サドバリー, チャーチ ストリート 34  
 (74) 代理人 弁理士 山本 秀策

(54) 【発明の名称】 自動翻訳および再翻訳システム

## (57) 【要約】

英語などのソース自然言語での入力テキストを、日本語などの標的自然言語での出力テキストへ、その入力テキストが発生されているときに翻訳するための自動自然言語翻訳システムは、文書またはファイルの中のソース言語の入力テキストの作成において一時休止が発生する場合を判定し、そしてその一時休止にตอบสนองしてその一時休止が検知された点まで、目的言語での入力テキストの翻訳を実行させる翻訳起動部を含む。この翻訳システムは、さらに、テキストが発生されている間に、特定のテキスト構造を検知したことにตอบสนองして、あるいは入力機能にตอบสนองして、翻訳を実行させ得る。自動翻訳が実行されるたびに、翻訳はそのテキストが入力された単語処理文書の先頭、即時受信された入力テキストの先頭、あるいはページまたはパラグラフの先頭などの開始点において翻訳を開始することができ、したがって、文章構造における任意の変化または入力テキストに対する最新の編集が、出力のテキスト翻訳に反映されることを可能にする。

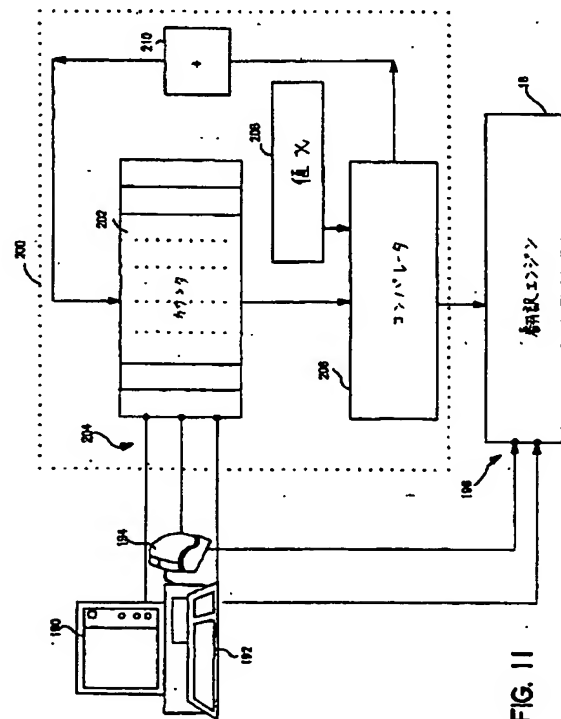


FIG. 11

## 【特許請求の範囲】

1. 自動自然言語翻訳システムであって、以下：

メモリ・モジュール、

該メモリ・モジュールと通信して、ソース自然言語での入力テキスト情報を受信し、そして該入力テキスト情報を該メモリ・モジュールの中に格納する、受信モジュール、

該受信モジュールと通信し、該メモリ・モジュールにアクセスし、そして該ソース自然言語での該入力テキスト情報を標的自然言語での出力テキスト情報に翻訳する、翻訳エンジン、ならびに、

該受信モジュールと通信し、該受信モジュールが所定の期間の間に入力テキスト情報を受信しなかった場合を判定し、そして該格納されている入力テキスト情報を翻訳するように該翻訳エンジンに対して命令信号を送信する、処理モジュール、

を含むシステム。

2. 請求項1に記載のシステムであって、前記ソース自然言語が英語であり、そして該標的自然言語が日本語である、システム。

3. 請求項1に記載のシステムであって、前記受信モジュールと通信し、前記ソース自然言語での入力テキスト情報を提供するユーザデバイスをさらに含む、システム。

4. 請求項3に記載のシステムであって、前記受信モジュールがユーザが前記ユーザデバイスを使って前記ソース自然言語での文書を作成しているときに、該ユーザデバイスから入力テキスト情報を受信するシステム。

5. 請求項3に記載のシステムであって、前記ユーザデバイスが、コンピュータ・キーボード、マウス、タッチスクリーン、または音声活性化送信機の中の

少なくとも1つを含む、システム。

6. 請求項5に記載のシステムであって、前記コンピュータ・キーボード上でエンター・キーの押下を要する信号を前記受信モジュールが受信したことの判定が行われる場合に、前記処理モジュールが前記指示信号を送信する、システム。

前記時間経過が存在する場合、前記ソース自然言語から前記標的自然言語へ前記文字ストリームを翻訳するステップ、

を含む方法。

16. 請求項15に記載の方法であって、以下の工程：

前記時間経過が所定の期間に等しいかどうかを判定するステップ、および

前記時間経過が前記所定の期間に等しい場合、前記文字ストリームの中の最初の文字から始めて、前記ソース自然言語から前記標的自然言語へ該文字ストリームを翻訳するステップ、

をさらに含む方法。

17. 請求項16に記載の方法であって、前記所定の期間が約0.3秒～2秒の範囲内にある、方法。

18. 請求項15に記載の方法であって、前記文字ストリームが前記ソース自然言語での少なくとも1つの単語を含む、方法。

19. 請求項15に記載の方法であって、前記文字ストリームが前記ソース自然言語での少なくとも1つの文を含む、方法。

20. 請求項15に記載の方法であって、以下の工程：

コンピュータ・キーボードを使って前記文字ストリームを発生するステップ、

前記ソース自然言語での該文字ストリームおよび前記標的自然言語での前記翻訳された文字ストリームをコンピュータ・モニタ上で表示するステップ、

をさらに含む方法。

21. 言語学習システムにおける自動自然言語翻訳システムであって、

ユーザワークステーションとインターフェースし、かつ該ユーザワークステーションにおいてユーザによって発生されたソース自然言語での文字ストリームを受信する受信モジュール、

該受信モジュールと通信し、該文字ストリームにおける文字の受信から、ある時間が経過したかどうかを判定する処理モジュール、

該処理モジュールと通信し、該ソース自然言語での該文字ストリームを該標的自然言語での文字ストリームへ翻訳する翻訳モジュール、ならびに

7. 請求項6に記載のシステムであって、前記受信モジュールが前記マウスのクリックを要する信号を受信したことの判定が行われる場合に、前記指示信号を前記処理モジュールが送信する、システム。

8. 請求項1に記載のシステムであって、前記所定の期間が0.3秒～1秒の範囲内にある、システム。

9. 請求項1に記載のシステムであって、前記翻訳エンジンと通信して、前記ソース自然言語での前記入力テキスト情報および前記標的自然言語での前記出力テキスト情報を表示するディスプレイをさらに含む、システム。

10. 請求項1に記載のシステムであって、前記処理モジュールが、カウント値が前記所定の期間に等しくなるまで、該カウント値を増加するためのカウンタを、さらに含む、システム。

11. 請求項10に記載のシステムであって、前記受信モジュールが入力テキスト情報を受信する場合に、前記カウンタがリセットされる、システム。

12. 請求項1に記載のシステムであって、前記翻訳エンジンが、前記指示信号の送信時に先立って、前記受信モジュールによって受信された入力テキスト情報を、翻訳する、システム。

13. 請求項1に記載のシステムであって、前記入力テキスト情報が入力デバイ

スから送信される文字のストリームを含む、システム。

14. 請求項1に記載のシステムであって、前記処理ユニットが、前記翻訳エンジンに対して前記指示信号を送信する前に、前記所定の期間に等しい、前記入力テキスト情報における時間経過の存在を判定することによって、前記受信モジュールがテキスト情報を受信していない場合を判定する、システム。

15. 文字ストリームの自然言語翻訳を自動的に実行する方法であって、

複数の文字を含んでいるソース自然言語での文字ストリームをユーザデバイスから受信するステップ、

前記文字ストリームの受信において時間経過が存在するかどうかを判定するステップであって、該時間経過の間に前記文字ストリームは受信されない、ステップ、

該翻訳された文字ストリームを該ユーザワークステーションへ送信する送信モジュール、

を含むシステム。

## 【発明の詳細な説明】

## 自動翻訳および再翻訳システム

## 関連出願および特許への相互参照

本願は、1996年4月23日、1996年6月14日、そして1997年6月9日それぞれ特許協力条約（PCT）を通して米国受理官庁に出願された、対象国を日本と米国とする国際特許出願PCT/US第96/05567号、PCT/US第96/10283号、およびPCT/US第97/10005号の一部継続出願であり、これらはすべて参考として本明細書において援用する。米国特許第5,528,491号もまた、関連特許であり、その内容全体を参考として本明細書において援用する。

## 技術分野

本発明は、自動自然言語翻訳に関し、特に、文書が、作者、ユーザまたはシステムによって生成されている時に、文書が自動的に翻訳される自然言語翻訳システムに関する。

## 背景情報

自然言語の機械ベースの翻訳については、種々の方式が提案されてきた。代表的には、翻訳に使用されるシステムは、ある自然言語での入力を受信し、そして操作を行って、別の自然言語での出力を供給するコンピュータを含む。このタイプの翻訳は不正確で長々しいものであり、そして得られた出力は熟練したオペレータによる大幅な編集を必要とすることが多い。

公知のシステムにより実施される翻訳操作は、一般に構造変換操作を含む。この構造変換の目的は、ソース言語の文の解析ツリー（すなわち、構文構造ツリー）を標的言語における、対応するツリーに変換することである。二つのタイプの構造変換方式（文法規則に基づく変換およびテンプレート間の変換）が試みられてきた。

文法規則に基づく構造変換において、構造変換の領域は、ソース言語の解析ツリー（すなわち、与えられたノードの直接の子であるサブノードの集合）を得る

ために使用されている文法規則の領域に限定される。例えば、以下の式：

、より難しい。

## 発明の要旨

本発明に従う自動自然言語翻訳システムは、公知の機械ベースの翻訳装置に対して多くの利点を有する。本発明のシステムが、自動的に入力テキスト情報の最良の可能な翻訳を選択し、そしてユーザに出力（例えば、英語の入力テキストの日本語訳）を提供した後で、ユーザは、表示された翻訳を編集するか、または自動的に他の翻訳を入手するために、このシステムと対話し得る。本発明の自動自然言語翻訳システムをオペレータは、より生産的であり得る。なぜなら、本システムによって、オペレータは、残りの部分を自動的に再翻訳させながら、これでいいと判断した翻訳の部分だけを保持することが可能になるからである。この選択的な再翻訳操作は、再翻訳を必要とする部分に対して正確に指向されるので、オペレータは、潜在的に多くの不正確ではあるが、非常に質の高い翻訳を検討する時間および退屈な仕事から解放される。さらに、このシステムは、翻訳調整において任意の精度を可能にするので、通常は、翻訳の最終構造の多くがシステムによって生成される。したがって、このシステムは、人間（オペレータ）のミスについての可能性を少なくし、構造、人称および時制の変更を含む編集における時間が節約する。このシステムはオペレータに広範囲でかつ信頼性のある文法および正字法の知識の利益を充分に与える。

本発明の自動自然言語翻訳システムの、ソース言語における曖昧な文区切りの、融通のきく処理、および強力な意味伝達、さらなる精度およびオペレータの翻訳の編集の減少を提供する。保存された統計学的情報はまた、特定のユーザサイトに対する好ましい翻訳を修正することにより、翻訳の精度は改善される。本システムの熟語処理方法は、熟語を構成する単語の列を偶発的に含んでいる文を、熟語の意味を解釈することなく、正確に訳すことを可能にする点で有利である。本システムは効率的であるが、なお長距離特性適合のような融通のきく機能を有する。本システムの構造バランスエキスパートおよび同格構造エキスパートは、目的とする解析と目的としない解析とを効率的に識別する。大文字エキスパートは、文中の大文字の単語の正確な解釈を効率的に入手し、そして大文字列手順は、

VP=VT01+NP（動詞句(VerbPhrase)は、この順序で目的語の一つの他動詞(SingleObjectTransitive Verb)および名詞句(NounPhrase)からなる。）

および

日本語：1+2=>2+1（VT01とNPとの順序を逆転させる。）

を仮定する。

規則の適用を含む各ソース言語の解析ツリーは、構造的には動詞と目的語との順序が反対になるように変換される。なぜなら、日本語では動詞はその目的語のあとに来るからである。この方法は、特定の交換が適用される場所を見つけることが容易であるという点で非常に効率的であり；ソース言語の解析ツリーを得るために規則が使用された場所で正確に適用される。他方、上に述べたように、その領域が大幅に制限され得るという点、および自然言語が、同胞ではないノードにまたがるような変換規則を必要とし得る場合があるという点で、弱い交換機構であり得る。

テンプレート間の構造変換では、構造変換は、入力/出力(I/O)テンプレートまたはサブツリーの形で指定される。ある所定の入力テンプレートがある所定の構造ツリーと一致する場合には、そのテンプレートと一致する構造ツリー部分が、対応する出力テンプレートによって指定されたとおりに変換される。これは非常に強力な変換機構であるが、所定の入力テンプレートが、所定の構造ツリーのいずれかの部分と一致するか否かを見つけるのに長時間かかり得るという点で、高価であり得る。

それにもかかわらず、そのタイプの変換が従来の翻訳システムによって実行され、そのようなシステムはユーザまたはオペレータがシステムへの文書のタイピングを完了した後、ユーザまたはオペレータからの直接の翻訳コマンドが存在する場合に、テキストの翻訳を容易にするだけである。結果として、これらのシステムは言語の学習を行わず、あるいはユーザまたはオペレータが単語ごと、または文ごとベースでその翻訳を調べることができず、したがって、異なる文脈に

おける単語の翻訳のバリエーションをユーザまたはオペレータが検出することは

普通名詞としての解釈を完全には無視しないで、複合語の固有名詞を効率的に処理する。

1つの局面において、本発明は、自動自然言語翻訳システムの改良に関する。ここで、この改良は、入力テキストが発生されているときに、文書またはファイルにおけるソース自然言語、例えば、英語を、標的言語、例えば、日本語、スペイン語、あるいは他の自然言語での出力テキストへと自動的に翻訳することに関する。1つの局面において、この自然言語翻訳システムは、例えば、入力文字のストリーム中の一時休止のような、入力テキストの入力時に一時休止が検知された時に翻訳が生成する。別の局面において、自然言語翻訳システムは、以前に翻訳された入力テキストに沿って、最近受信した入力文字ストリームを自動的に翻訳し得、それによってその入力テキストの特定部の再翻訳を生成し得る。別の局面において、この自然言語翻訳システムは、以前に受信した入力テキストに対する修正または変更を生じた入力テキストを自動的に翻訳し得る。

本発明の別の局面において、自動翻訳は、入力テキストにおける開始点（たとえば、文書、パラグラフ、単語の先頭または他の位置）において、開始し得、したがって、文構造における任意の変更、または編集を最終の翻訳の中に吸収させかつ反映させ得る。

本発明の別の局面において、自動翻訳は、以前に受信した入力テキストに対する修正が行われたその入力テキストの一点（たとえば、以前に受信されているが、最近編集されたパラグラフの途中）において、開始し得る。

別の局面において、本発明はグローバルな自然言語の自動翻訳システムの改善に関し、ここで、その改善はソース自然言語（たとえば、英語）での入力テキストを、たとえば、文の断片などのある種のテキスト構造が発生した後、標的言語（たとえば、日本語、スペイン語または他の自然言語）での出力テキストに自動的に翻訳および再翻訳することに関する。

別の局面において、本発明は、自然言語自動翻訳システムの改善に関し、ここで、この改善は、ユーザ、オペレータ、またはシステムが、タッチスクリーンまたはキーボード上のエンター・キーの押下げ、アイコン上でのマウスのクリック、あるいは他の入力信号（たとえば、文字ストリームを発生しない機能信号）を

伝

違する信号を送信する場合に、ソース自然言語（たとえば、英語）での入力テキストを、標的自然言語（たとえば、日本語、スペイン語または他の自然言語）での出力テキストに自動的に翻訳および再翻訳することに関する。

さらに別の局面においては、本発明の自動自然言語翻訳システムは、一時休止が検知される場合にカウントを開始するカウンタを使うことによって、入力 of 文字ストリームにおける一時休止にตอบสนองして、自動的に翻訳を実行し得る。本発明の別の実施態様において、本発明の自然言語の自動翻訳システムは、ある種のテキスト構造、またはシステム機能に関連している入力信号が検知される場合に、自動的に翻訳を実行し得る。

本発明の別の局面において、本発明の自動翻訳システムは、入力テキストが生成されているときに直ちにその入力テキストを翻訳すること、およびその翻訳が生成しているときにその入力テキストと並べて、翻訳された出力テキストを表示することによって、言語学習ツールとして使用され得る。

本発明の上記および他の目的、局面、特徴および利点は、以下の説明と請求の範囲から、より明らかにになる。

#### 図面の簡単な説明

図面中、同じ参照文字は、全体にわたり、別の図面であっても同じ部分を示す。また、図面は必ずしも同一の縮尺ではなく、その代わり一般的に本発明の原理を例示する際に強調が施される。

図1は、自然言語の自動翻訳を行うためのシステムを示すブロック図である。

図2は、図1のシステムの全体的な機能を示すデータフローチャートである。

図3は、図1のシステムの操作を示すフローチャートである。

図4は、図1のシステムのプリパーサの文末認定機能の操作を示すフローチャートである。

図5は、図1のシステムのパーサの操作を示すフローチャートである。

図6は、図1のシステムの意味伝達の操作を示すフローチャートである。

図7は、図1のシステムの構造変換の操作を示すフローチャートである。

さらに他の実施態様において、本発明のシステムは英語からスペイン語へ、およびスペイン語から英語へ翻訳し得る。さらに他の実施態様において、本発明のシステムは、英語から複数の異なる言語へ、あるいは英語でない言語から英語でない別の言語へ翻訳し得る。

本発明のシステムは、ソース自然言語を受信および記憶するための手段；標的自然言語への翻訳を生成する翻訳エンジン；ユーザに翻訳を表示するための手段；およびユーザのために別の翻訳を得、そしてユーザに対して表示するための手段を含む。本発明のシステムの実施態様において、翻訳エンジンは、翻訳起動部、プリパーサ、パーサ、グラフ作成部、評価部、グラフ・スコアラ、解析抽出部および構造変換部を含む。翻訳起動部は、いつでもおよび/またはどんな状態において翻訳が実行されるべきかを決定する。プリパーサは、入力テキストを検査し、そして入力文の境界における任意の曖昧さを解消する。その後、プリパーサは、辞書の見出し語を用いて選択された(seed)解析チャートにおいて入力テキストを生成させ、および表示する。パーサは、上記チャートを解析して、入力テキストに対する可能な構文の分類を入手する。グラフ作成部は、解析チャートに基づいて、入力テキストの可能な構文解釈のグラフを作成する。このグラフには、入力テキストの可能な解釈に関連するノードおよびサブノードが含まれる。一連のエキスパートを含む評価部は、可能な解釈のグラフを評価し、そしてグラフのノードおよびサブノードにエキスパートの重みを加える。グラフ・スコアラは、サブノードをスコア付けするためにエキスパートの重みを使用し、その後でこのグラフ・スコアラは、N個の上位のスコアと各ノードとを関連づける。解析抽出部は、グラフ・スコアラによって決定された好ましい解釈に対して、1つの解析ツリー構造を割り当てる。構造変換部は、標的自然言語での翻訳を入手するために、解析ツリー構造に関する構造変換操作を行う。

以下の三つの段落では、(a) 各サブノードに対する最終重み付けスコアを計算するために、グラフ・スコアラがどのようにエキスパートの重みを組み合わせるか；(b) グラフ・スコアラがどのようにサブノード・スコアを組み合わせる最終ノード・スコアに到達させるか；(c) 言語学的情報がノードおよびサブノード

図8は、図1のシステムのエキスパート評価部のフローチャートである。

図9は、例示としての句「by the bank」について、図1のシステムが使用する

サンプル・グラフの図である。

図10Aは、本発明の1つの実施態様に従って、入力テキストを自動的に翻訳するためのシステムを示す機能的ブロック図である。

図10Bは、本発明の別の実施態様に従って、入力テキストを自動的に翻訳するためのシステムを示す機能的ブロック図である。

図11は、本発明による翻訳起動部の1つの実施態様を示すブロック図である。

図12Aは、本発明の1つの実施態様による翻訳起動部の操作を示すフローチャートである。

図12Bは、本発明の他の実施態様による翻訳起動部の操作を示すフローチャートである。

図13A、図13B、および図13Cは、本発明のシステムを使用して、ソース言語から標的自然言語への翻訳を示す画面表示の一例を示す。

図14は、本発明の別の実施態様による翻訳起動部の操作を示すフローチャートである。

図15は、本発明の別の実施態様による翻訳起動部の操作を示すフローチャートである。

図16は、本発明の別の実施態様による翻訳起動部の操作を示すフローチャートである。

#### 詳細な説明

最初に、図面を何ら参照しないで、本発明による自動自然言語翻訳システムの一般的記載を提供する。本発明のシステムの一般的記載の後で、種々の図面を参照する。

本発明の自動自然言語翻訳システムは、ソース自然言語を標的自然言語に翻訳し得る。1つの実施態様において、本発明のシステムは英語を日本語に翻訳し得る。別の実施態様において、本発明のシステムは、日本語を英語に翻訳し得る。

ードのツリーを介してどのように伝達されるか、についての説明を提供する。

各サブノードに対する最終重み付けスコアを計算するために、グラフ・スコアラは、各サブノードに定数値を関連づける。各サブノードに関連する言語学的情報の分析により、サブノードのスコアが決定される。例えば、一連のエキスパート評価部が、各ノードおよびサブノードに記憶された言語学的情報を検査する図

8を参照されたい。グラフ・スコアラは、特定のノードまたはサブノードに対する最終加重平均を入手するために、各エキスパートに対する個々の重み付けスコアを合計する。複数の重み付けスコアを一つの加重平均スコアに合わせることは、コンピュータ科学における標準的な課題である。使用され得る一つの方法としては、各エキスパートの結果に、そのエキスパートに割り当てられた定数(重み)を掛け合わせる方法がある。各エキスパートに割り当てられる重みは、設計での選択事項である。設計者は各エキスパートに割り当てる優先度(重み)を選択し得る。加重平均は、各数字に定数を掛け、一連の数字を合計したものである。例えば、以下の式：

$$\text{加重平均} = (w_1)(x_1) + (w_2)(x_2) + \dots + (w_n)(x_n)$$

ここで、重み $w_1, w_2, \dots, w_n$ は、いずれも負でない数で、かつ合計は1である。例えば、統計の期待値に関する加重平均の使用について記載される Spiegel, Theory and Problems of Probability and Statistics 76 (McGraw-Hill, Inc. 1975)を参照されたい。

サブノード・スコアを合わせて最終ノード・スコアを入手するために、グラフ・スコアラは、グラフの底辺の部分から一番上の部分にサブノード・スコアを伝達し得る。各ノードがN個のスコアの集合を有するグラフの場合には、一つまたは複数の伝達方法を決定することが可能である。サブノード・スコアを伝達するのに使用され得る一つの技術としては、最適化問題を解くために使用されるある型の動的プログラミングである記憶方法がある。最適化問題の解法は、多くの可能な数値(結果)を含み得る。課題は最適値を発見することである。最適化に使用するアルゴリズムは、各サブサブ問題を一回だけ解き、そして結果を記憶するので、サブサブ問題に遭遇するごとに答を再度計算する必要がなくなる。最適化

問題に適用されるより詳細な説明については、例えば、Cormenら、Introduction to Algorithms 301-314 (Mcgraw-Hill Book Co, 1990) を参照されたい。Introduction to Algorithmsの301、302および312ページに記載される方法は、グラフを介してサブノード・スコア情報を伝達するために使用さ

れ得る一つの方法である。

言語学的情報をツリーを介して伝達する場合には、システムの意味伝達部分は、構成要素を含む意味情報をより小さな構成要素からより大きな構成要素へと伝達するように操作する。意味伝達は、解析操作で使用する統語分類の四つのクラス (SEMP, SEMP, SEMADJおよびVERB) に適用される。意味伝達が行われ得る前に、ノードに記憶されている言語学的情報を分析しなければならない。ノードに記憶されている意味情報の分析は、文法規則において、名詞様構成要素と動詞様構成要素を点検することにより、どの名詞様目的語に、動詞様構成要素のどの選択的制限スロットを適用するかを識別する規則の集合によって導かれる。Gerald Gazdarは、その著書Natural Language Processing In Prolog (Addison-Wesley Publishing Co, 1989) で、本明細書において開示されるグラフに類似する、配向性非環状グラフにおけるノードに記憶されている意味情報を分析するのに使用され得る規則の集合について説明している。Gazdarは、隣接するノードに関する情報と一致させる特性マッチングの使用について説明している。Gazdarは、特性マッチングは、以下のことを記述する式を含むと述べている。

あるノード上に現れる特定の特性は、別のノード上に現れる特性と同じものであるに違いない。最新の研究は、親の分類とそれらの特性に関連する形態素を示す子上に現れるあるクラスの特性の詳細を等しくすることを担う原則を仮定している。この子は、その句の「頭」と呼ばれる。大部分の句は頭を一つだけ有する。従って、例えば、動詞句は、その動詞の時制を受け継ぐ。何故なら、その動詞は動詞句の「頭」であるからである。これまで本発明者らが使用してきた表記法の資源を使用しても、文法全体を基礎とするこの原則を特定する簡単な方法はない。しかし、本研究者らが、関連する特性がDAGの一本の枝上にすべて発見すると仮定する場合、規則ごとに非常に簡単にこの原理の効果を規定し得る。次の

「は、代表的に名詞句 (NP) と分類される。いくつかのモデルはまた、「an arrow」を構文上の特性 (例えば、これが単数の名詞句であること) および意味特性 (これが武器に言及する) を用いて分類する。句が曖昧である場合、この句を分類する2つ以上の方法がある。例えば、「an arrow」はまた、矢の形をした

記号にお言及し得る。言語学的モデルが、曖昧さを解消するための方法を提供する場合、これらのモデルは通常、より小さな単位をより大きな単位に結合することによって曖昧さを解消する。より大きな単位を評価する場合、これらのモデルは、より大きい単位に含まれている情報の部分だけを考慮する。

本発明のシステムの例示の実施態様において、「an arrow」(記号対武器)の意味上の特性が、「They would like an arrow.」という文の「like an arrow」という動詞句を評価する際に使用される。対照的に、「an arrow」という句の構文が「He shot it with an arrow.」におけるように存在した場合には、「an arrow」の意味特性は、「shot it with an arrow.」という動詞句を評価する際には使用されない。

特定の言語学的モデルを単一の方法で解釈した文の任意のサブストリング (解釈したサブストリング) について、エクスポートされた属性が存在する。エクスポートされた属性は、解釈したサブストリングと、より大きなサブストリングを形成する他の単位との組み合わせを評価するために使用されるすべての属性である。エクスポートは、エクスポートされた属性と一緒に解釈される、解釈したサブストリングである。解釈したサブストリングに含まれているが、エクスポートされていない属性は、サブ構造と呼ばれる。

本発明のシステムのパーサは、文法データベースを含む。パーサは、文法規則を使用して文のすべての可能な解釈を見出す。文法データベースは、 $X = A_1 A_2 \dots A_n$  の形の、一連の文脈自由句構造規則からなる。Xは、 $A_1, A_2, \dots, A_n$  から構成または形成されていて、そしてより低いノード (サブノード)  $A_1$  から  $A_n$  のより高いノードと呼ばれる。

本発明のシステムのグラフ作成部は、一つの文に関して可能な多くの解釈を図形で表す。グラフの各ノードは、いくつかのサブストリングのエクスポートに対

で、本研究者らは、代表的なVP規則を以下のように書き表し得る。

$VP \rightarrow V NP PP$

$\langle V \text{の頭} \rangle = \langle VP \text{の頭} \rangle$ 。

これは、Vの「頭」の特性値と、親であるVP上の「頭」の特性値とが同じものであることを要求する。」

のであることを要求する。」

Gazdarで議論されている規則は、本明細書に記載する構文の各分類に容易に適用され得る。Gazdarの規則を使用して各ノードに割り当てられた言語学的情報は、記憶方法の技術を用いて、ツリーを介して伝達され得る。

従って、これまでの三つの段落の内容を要約すると、加重平均は、サブノード・スコアを決定する方法であり、各サブノード・スコアは、最適化問題に適用した公知の記憶方法技術を使用してグラフを介して伝達され得、そしてGazdarの著書に述べられている戦略は、各ノードに記憶されている言語学的情報を分析するのに使用され得、そしてこの言語学的情報は、記憶方法技術を使用して解析ツリーチャートを通じて伝達され得る。

自動自然言語翻訳システムは、最初の自動翻訳の後に、自動的再翻訳を実行し得る。すなわち、システムが自動的に入力テキスト情報の可能で最も適切な翻訳を選択し、そしてユーザに出力 (好適には、英語の入力テキストの日本語の翻訳、または日本語から英語への翻訳) を提供した後、次いで、ユーザは表示されている翻訳を編集するために、または自動的に別の翻訳を入手するために、システムと対話し得る。

自動自然言語翻訳システムは、一つの文をサブストリングに分解する言語学的モデルを使用する。サブストリングは、その文の一部として、特定された順序で出現する一つまたは複数の単語である。例えば、サブストリング「The man is happy」は、「The」、「The man」、「man is happy」、「is」および「The man is happy」それ自身を含むが、「is man」、「man man」、「The is」は含まない。

異なる言語学的モデルは、種々の方法および種々の詳細レベルでサブストリングを分類する。例えば、「They would like an arrow」においては、「an arrow

応する。本発明のシステムの実施態様において、一つのエクスポートは一つのノードで表される。グラフは、一つのエクスポートに関連するノードから出ている円弧を含む。円弧は、文法規則の適用に基づくエクスポートのサブ構造を表す。グラフは少なくとも二つのタイプの円弧を示す: (1) 同じサブストリングの一つの異なるエクスポートを指す単一の円弧、(2) 二つのエクスポートを指す一

組のポイントを含む二成分の円弧、ここで、連結されるときは、そのサブストリングは、もとのエクスポートのサブストリングを形成する。(2)に記載される公式は、チョムスキーの正規形態における文法を仮定していることに留意された。

グラフはまた、一つの出発エクスポートSを含み、そこから一連の円弧を辿ることによりグラフのすべての部分に到達し得る。出発エクスポートは、文全体に対応する。

同じエクスポートが1以上のエクスポートから構成され得る場合に限り、一つのノードから複数の円弧が出る (二成分の円弧における一対のポイントは、この目的では複数の円弧とは考えない)。そのエクスポートが複数のエクスポートの一つの成分である場合に限り、複数の円弧が一つのノードを指す。円弧が出ていないノードは、サブストリングに割り当てられた辞書の見出しに対応する。

複数の言語学的エクスパートが、エクスポートの集合に数値スコアを割り当てる。この言語学的エクスパートは、グラフの各ノードに上記スコアを適用する。本発明のシステムの実施態様において、スコア付けアレイ (ここで、アレイの各要素は、特定のエクスパートのスコアに掛け算を行うための重みである) は、任意の所定の文に対する浮動小数点数の固定の長さ「N」である。

上記スコアは、グラフ作成エンジンおよび/またはパーサに組み込まれるスコア付けモジュールにより評価される。スコアは、より高いエクスポートを形成しているすべてのエクスポートに対して計算される。より高いエクスポートに対するスコアは、よりレベルの高いエクスポートを形成しているエクスポートと構造バランスエクスパートによって割り当てられるスコアのような、組み合わせに適用される任意のエクスパートのスコアとの合計として計算される。

ノードが訪問されそしてスコア付けされる順序は、標準的な深さ優先グラフ移動アルゴリズムである。このアルゴリズムでは、スコア付けされたノードにはマークが付けられ、そして再びスコア付けられることはない。スコア付けのプロセス中、スコア付けモジュールは、より高い任意の単位ノードの評価を行う前に、辞書の見出し語ノードを評価する。辞書の各見出しは一つのスコアを生じる。

一つのエクスポートを作成するのに複数の方法がある場合には、複数のスコア

が生じる。すなわち、エクスポートを作成するのにkとおりの方法は、k個の可能なスコアを生じる。複数のスコアは以下のように処理される。

(1) 一成分規則では、より低いエクスポートのk個の各スコアが、一成分規則に適用されるエクスポートの数値に加算され、そして得られたk個のスコアのベクトルは親のエクスポートに関連付けられる。

(2) 二成分規則では、左の子はgスコアを有し、そして、右の子はhスコアを有するものと仮定する。次いで、hスコアのg倍の合計が、左の子の各スコアに右の子の各スコアを加算し、さらに、二成分規則に適用されるエクスポートの数値を加算することによって計算される。hのg倍がNを超える場合には、N個の最良のスコアだけが親のノードと一緒に保持される。

(3) ノードのエクスポートが複数の方法で作成され得る場合には、たかだかN個のスコアがそのノードのスコア・リストに加算され、最良のスコアが保持される。

スコア付けが完了すると、上記方法は、各エクスポート、そのノードと、そのエクスポートに表示されていないすべてのサブ構造属性を含む、エクスポートを作成するためのg個の最も可能性の高い方法(言語学的モデルに関する)を表すg個のスコア(1からNまでの範囲内のg)の集合とを関連づけたことを確認する。ルート・ノードSの特定の場合では、このスコア付け方法は文を形成するg個の最も可能性の高い方法を生じる。

上記に記載される各スコア・リストの各スコアは関連するポインタを有する。ポインタは、より低いエクスポートのスコア・リストのどのスコアが、より高いレベルのスコアを作るために組み合わせられたのかを示す情報を提供する。各ポイン

ワードおよび文法規則を含む、翻訳を行うために使用されるかまたは有用であるすべての種類の情報を記憶するのに使用される。ユーザ入力インターフェース22は、キーボード、マウス、タッチスクリーン、ライトペンまたは他のユーザ入力デバイスを含む、本発明のシステムのオペレータにより使用される。ディスプレイは、コンピュータ・ディスプレイ、プリンタまたは他のタイプのディスプレイであり得、または、これは、オペレータに情報を伝達する他の手段を含み得る。出力インターフェース14は、日本語のような標的言語でのソース・テキストの最終翻訳を、伝達する。上記出力インターフェース14は、プリンタ、ディスプレイ、音声インターフェース、モデムまたはシリアルラインのような電子インタ

ーフェースを含み得、あるいは、末端ユーザにそのテキストを伝達するための他の手段を含み得る。

本発明の翻訳システムの1つの実施態様の操作において、図1を参照して、翻訳エンジン16は翻訳起動部21を含み得る。翻訳起動部は翻訳エンジン16の内部でこの図の中に示されている残りの要素と一緒に構成され得る。あるいは、図11に示されているように、翻訳エンジン16の他の要素と定期的にインターフェースし得る別のコンポーネントとして構成され得る。翻訳起動部21は入力テキストの翻訳をいつ開始すべきかを決定するためのコンピュータ処理ユニットを含み得る。翻訳起動部21は、代表的にこの決定を支配するパラメータを設定し、テキストが発生されている間に、そのようなパラメータに関係する状態が存在しているかどうかを判定し、そしてそのような状態が存在しているときに自動翻訳を起動する。本明細書における例示的な実施態様においてさらに説明されるように、翻訳起動部21は入力テキストの発生における時間経過がいつ存在するかを判定するために使用され得る期間または値が発生してそして格納し得る。翻訳起動部は、そのような期間に実質的に等しい時間経過または一時休止を認識する際に、本明細書で記載されている翻訳エンジン16において実行され得る自動翻訳および/または再翻訳を起動する。同様に、翻訳起動部は、入力テキストを処理し、そして単語、文または文の断片、およびパラグラフの形成を、上記の各種

タを追跡することにより、その文のg個の最も可能性の高い解釈を、曖昧でない解析ツリーとして抽出し得る。

今度は図1-図9を参照しながら、自動自然言語翻訳システムをさらに詳細に開示する。その後、図10-図16を参照しながら、本発明従う種々の改良点について記載する。

図1および図2を参照すると、本発明に従う自動自然言語翻訳システム10は、入力インターフェース12、翻訳エンジン16、記憶部18、ユーザ入力デバイス

22、ディスプレイ20および出力インターフェース14を含む。入力インターフェース12は、例えば英語または日本語のようなソース言語でのテキストの列を受信するように構築される。入力インターフェース12は、キーボード、音声インターフェースまたはモデムもしくはシリアル入力のようなデジタル電子インターフェースを含み得る。翻訳エンジン16は、記憶部のデータとともに、ソーステキストに対して翻訳操作を行う。翻訳エンジン16は、全体がハード配線の論理回路から構成され得、あるいは一つまたは複数の処理ユニットおよび関連する記憶命令を含み得る。

翻訳エンジン16は、以下の要素またはその部分を含み得る。すなわち、翻訳起動部21、プリパーサ24、パーサ26、グラフ作成部28、解析/翻訳評価部30、解析抽出部32、構造変換部34、および別の解析システム37を含むユーザ・インターフェース42である。構造変換部は、文法規則制御による構造変換部36、辞書制御による構造変換部38および生成規則制御による構造変換部40を含み得る。記憶部18は、ディスク(例えば、ハードディスク、フロッピーディスクおよび/または光学的ディスク)および/またはメモリ(例えば、RAM)などの記憶部の一つまたは複数の領域を含み得る。記憶部は、以下の要素の全部または一部を記憶し得る。基本辞書44、技術用語辞書46、ユーザが作成した辞書、文法規則48、生成規則50、意味特性ツリー52、構造ツリー54およびグラフ56。記憶部18は、またソース自然言語での入力テキスト情報、標的自然言語での出力テキスト情報、および一つまたは複数の辞書、領域キ

造を表すように設定された所定のパラメータを使って、検知し、そしてそのような構造が存在するときに翻訳を実行する。翻訳起動部およびその操作は、本明細書において図10-図16においてさらに説明される。

プリパーサ24は、まずソース・テキスト23に対して予備解析操作(ステップ102)を行う。これは、図10-図16にさらに示すように、プリパーサ24は、コンピュータ・ファイルまたは文書において作成、編集、受信またはそうでなければコンパイルされるときにソーステキストと含み得る。この操作には、ソース・テキストにおける文境界の曖昧性の解析が含まれ、そして辞書の見出し25を用いて選択された解析チャートを作成する。次いでパーサ26は、構文の可能性27で満たされた構造解析チャートを入手するために、プリパーサが作成した上記チャートの解析を行う(ステップ104)。グラフ作成部28は、解析

ステップから得られた解析チャートに基づいて、可能な解釈29のグラフを作成する(ステップ106)。一連のエクスパート43にアクセスする評価部30は、記憶された解釈のグラフを評価し(ステップ108)、そしてグラフ31にエクスパートの重みを加える。グラフ・スコアラ33は、ノードのスコアをつけ、N個の(例えば、20個の)最良のスコアをそれぞれに関連づける35。解析抽出部32は、この好適な解釈に解析ツリー構造39を割り当てる(ステップ110)。次いで、変換テーブル58にアクセスする構造変換部34は、標的言語での翻訳41を入手するために、上記ツリー上で構造変換操作を行う(ステップ112)。ユーザは、別の翻訳を入手するために別の解析システム37と通信し得る。

図4を参照すると、本発明のシステムは、入力ストリームを個々の句読点および語を形成する文字グループを含むトークンに分割することによって、予備解析操作を行う(ステップ114)。スペースの出現は、このレベルでの文字の解釈に影響を与える。例えば、「x-y」の「-」はダッシュであるが、「x-y」の「-」はハイフンである。

次いで、プリパーサは上記のトークンを単語に組み合わせる(ステップ116)。このレベルで、プリパーサは、特殊な構造(例えば、インターネット・アド



レス、電話番号および社会保険番号)を一つの単位として認識する。プリパーサはまた、グループ分けのために辞書参照を使用する。例えば、「re-enact」が辞書に「reenact」として存在する場合は、文中で一単語となるが、存在しない場合は、別々の三つの「単語」のままである。

次の予備解析段階は、文が何処で終わるかを決定することが含む(ステップ118)。この操作の間、プリパーサは、各文の可能な文末箇所(すなわち、ソーステキストの各単語の後)について一連のステップに従って、基本辞書および技術用語辞書ならびに任意のユーザ作成の辞書にアクセスする。プリパーサは、示された特定の順にこれらのステップを行う必要はなく、そしてこれらステップは一連の順序付けられた規則として実行され得、またはハードウェアに組み込んでコード化され得る。

図5について参照すると、プリパーサは、ダッシュ記号の連続「——」のよ

うな解析できない任意の文字列をそれ自体を「文」として解釈し、そして記録するが、翻訳される文としては解釈も記録もされない(ステップ120)。プリパーサはまた、文の末尾であるためには1つの行において任意の列の二つのキャリッジ・リターンを要求する(ステップ122)。次の単語の最初の文字が小文字である場合には、プリパーサは一つの文の終わりで見なさない(ステップ124)。文が新たな行で始まりかつ短い場合には、プリパーサは、それをそれ自体で一つの文と見なす(例えば、タイトル)。

プリパーサは、閉じ括弧および閉じ引用符を含む特定の除いて、終止符、疑問符、または感嘆符を文の終わりで見なす(ステップ128)。「.」「」または

「?」等で終わっている文の場合には、プリパーサは、引用符の前の句読点に加えて、引用符の後に仮想の句読点を使用する。「?」を必要とする基本的な句読点についての別の例を、以下の例に示す。

The question was "What do you want?".

Did he ask the question "What do you want"??

Are you concerned about "the other people"?

拡張するために辞書を使用して単語の異なる意味を含ませ、そしてその文のすべての意味解釈を表す配向性非環状グラフを作成する。このグラフは、以下に説明する一連の意味伝達の手順の助けを借りて作成される。これらの手順は、一連の作成された文法規則に基づいて行われ、いくつかの場合には、意味情報についての意味特性ツリーにアクセスする。意味特性ツリーは、意味分類を含むツリー構造である。このツリーは、おおざっぱに抽象物から具体物へと組織されており、そしてツリーの中でのそれらの分離およびツリーのなかでのそのレベルの両方に関して、一対の用語が意味の上でどのように関連しているかをその手順が決定することを可能にする。例えば、「cat」および「dog」は、「cat」と「pudding」より関連性が高い。したがって、「cat」および「dog」の対は、ツリーの中ではより短い距離で分離している。「animal」および「cat」は、ツリーの異なるレベルに記憶されている単語の例である。何故なら「animal」は「cat」に比べて、より抽象的な用語であるからである。

図9について説明すると、このグラフは、ノード80および種々のタイプの関連を示す方法でポインタ88、89、90、91でリンクされたそのサブノード82、84、86を含む。グラフの第一の関連のタイプは、句を表すノードが、それを構成している単語のノードやサブフレーズのノードを指すポインタを有するものである。例えば、句「the bank」を表すノード84は、ポインタ92、93により、それを構成している単語「the」94および「bank」95にリンクされている。グラフの第二の関連のタイプは、句の解釈が、より低いレベルのものから、より高いレベルの同じ成分を作るもう一つの方法を指すポインタを有している場合である。例えば、句「by the bank」を表すノード80は、それぞれを構成する各単語にリンクしているポインタ88と89、および90と91を各々含む二つのソース解釈位置81、83を有し得る。この例の場合、異なる個々の構成語は、それぞれが句「the bank」に対して異なる意味を表す異なるサブノード84、86を含む。グラフの構造は、解析操作の結果により規定され、ソース文の構文により制限を受ける。このグラフのノードは、意味伝達のプロセス中に記入され得る意味情報に対する記憶位置と関連づけられる。

英語では、これらの各文は「?」で終わっている可能性が高い。プリパーサが追加した仮想の句読点は、引用符の前に疑問符であり得る何かが存在するのか、または全然何もないかのいずれかを示している。引用符の後ろには、終止符または疑問符であり得る何かが存在している。文の残りの文法的構造によって、その後の処理段階で最も適切な選択を行うことが可能となる。

プリパーサはまた、終止符の予備解析で、さらにいくつかのアプローチを使用する(ステップ130、132、134、136および138)。辞書に含まれるいくつかの短縮形には、文頭には決してこないというマークが付けられ、そしてその他には文末には決してこないというマークが付けられる(ステップ130)。これらの規則は常に尊重される。例えば、「Ltd.」は決して文頭にはこないし、「Mr」は決して文の終わりにはこない。プリパーサはまた、次の単語が「the」、「in」などのような頻出する文法的単語でないかぎり(ステップ13

2)、一字の大文字に続いて終止符を伴う文は終結させない。終止符の前の単語が、いずれかの辞書に載っている場合には、終止符のところで文は終わる(ステップ134)。終止符の前の単語がこの辞書に載っておらず、そしてその単語がその内部に終止符を有し(例えば、I.B.M.)、かつ次の単語が小文字形態では辞書に載っていないか、または次の単語そのものが大文字である場合には、これは文末ではない(ステップ136)。残りの場合には、終止符は文末のマークを付ける(ステップ138)。

再び図2および図3を参照すると、一旦、文の境界がプリパーサにより規定されると、パーサはその文の単語を構文分類に分け、そして、その文の可能な構文の解釈25を計算するために、それらの単語に文法データベースからの文法規則を適用する(ステップ104)。これらの文法規則48は、その言語の文法的な制限を表すコンピュータ読み出し可能な一連の規則として実行され得る。英語について、このような規則は何百存在し得、これら規則は何百の構文分類に適用され得る。この操作の計算に要する処理量を減らすために、一つの単語の異なる可能な意味は無視される。

次のステップ(ステップ106)において、グラフ作成部は、パーサの結果を

システムの意味を伝達する部分は、それらが含むより小さな成分からより大きな成分へと意味情報を伝達する働きをする。意味情報は、初めの段階の解析操作で使用される構文分類の四つのクラスに適用される。四つのクラスは、SEMP (名詞型の目的語や前置詞句を含む)、SEMP (通常主語を取る、目的語のような動詞句)、SEMDJ (形容詞)、およびVERB (多くの場合目的語を取る辞書の動詞型の動詞)である。その他の構文分類はある規則の中では無視される。文法規則の設定者は、規則に特定のマークを付けることによって、表面上に出ない行動も無効にすることもできる。これらの特別命令は最初にくる。

システム内を意味特性が伝達する様式には二つの局面がある。第一の局面は、文法規則の名詞型および動詞型の構成部分を検査することによって、どの名詞型の目的語に、動詞型の構成要素のどの選択的な制限スロットを適用するかを決める規則の集合である。例えば、「I persuaded him to go.」という文の動詞句に対する規則は、おおまかにいうと、 $VP = VT11 + NP + VP$  (ここで、 $VP$ は動詞句、 $VT11$ は他動詞の分類、 $NP$ は名詞句)である。例示としてのデフ

ォルト規則は、動詞が目的語を取る場合には、動詞の右側にある最初の $NP$ に、選択制限を適用すべきであることを示す。別の規則では、その主語に対する $VP$ の制限は、 $VP$ の左側に見出される最初の $NP$ に適用されるべきであると定めている。これらの規則は合わせて、「persuade him」および「him go」の両方の意味がよく通るように評価されることを確実にしている。すでに説明したように、これらの規則は英語の複雑な文法を反映していて、そのためにその数は非常に限定されている。

図6について説明すると、意味伝達操作は、選択制限をSEMPから命令文にコピーする操作を含む(ステップ140)。SEMPが位置の表現として使用されている場合には、良い位置を規定する意味定数に対してその妥当性が評価される(ステップ142)。ある規則が二つのSEMPの結合(構文上の特徴を論理積することによって検出された)を含んでいる場合には、グラフ作成部は、意味特性を論理積して、意味上の距離エキスパートに適用する(ステップ144)。

意味特性の伝達に対して特定された規則を検査している中で、グラフ作成部が



、もっと高いレベル（例えば、それがもっと多くの単語を含むSEMNPの一部となる）へ伝達するSEMNPの「頭」の位置を見つけた場合には、グラフ作成部は、意味特性をも伝達する（ステップ146）。しかし、「頭」が、区分用の単語（例えば、「Portion」、「part」である場合には、「頭」はかわりにSEMNPから左または右に伝達する。SEMNPおよびSEMADJが、区分用の位置を持っていない（ステップ148）場合だけを除いて、同じ方法で伝達される。形容詞は、この目的のためには、SEMNPのクラスの一部である。

SEMNPがVERBを含む規則から作られた場合は、グラフ作成部は、もしVPが受動態でなければ、VERBの主語制限を上の方に伝達させる。受動態の場合、VERBの最初の目的語制限がかわりに伝達される（ステップ150）。SEMNPを含む規則の場合には、グラフ作成部は、SEMNPの選択制限をSEMNPから左に移動するときには通過するNPに適用しようとする（ステップ152）。SEMADJを含む規則の場合には、グラフ作成部は、SEMADJの選択制限を、最初にSEMADJから右に移動するときには通過する任意のSEMNPに適用しようとし、それがうまくいかない場合には、左の方向に移動しようとする（ステップ154）。

VERBの任意の残りの未使用の目的語選択制限（受動態であるために上の方にそれまで伝達していない）に対して、グラフ作成部は、VERBの右側に通過するSEMNPに、上記の制限を順番に適用する（ステップ156）。これらすべての規則で、動詞選択制限は、それが何かに適用されるとすぐに使い尽くされてしまう。これまでのすべての規則の場合、SEMNPは、それらSEMNPに何かが適用された場合には、使い果たされない。この規則からスタートすると、SEMNPは使い尽くされる。最終的に、ある規則がSEMNPを作った場合に、グラフ作成部は、それまで使用されていないSEMNPまたはSEMADJが含まれているかどうかを決定し、含まれている場合には、それを上の方に向かって伝達する（ステップ158）。

システムはまた、言語特性の特性マッチングを行う。言語特性は単語および他の構成要素の特性である。構文の特性マッチングはパーサが使用し、意味特性マッチングは、グラフ作成部が使用する。しかし、同一の技術が両者に対して使用される。例えば、「they」は複数という構文特性を有し、一方「he」は単数とい

優先：long ago

普通：black sheep

オールマイティ熟語は、その列を構成している任意の単語の任意の他の可能な解釈よりも優先する。優先熟語は、同じ一般的なタイプの他の構成語であって、全く同じ単語を使用するものよりも優先する。普通熟語は、同じレベルで他の見出し語と競合する。

その結果得られるグラフは、エキスパート（ステップ108、図3）によって評価され、上記エキスパートはグラフの解釈の正確さの可能性を表すスコアを提供する。本発明のシステムは、文全体のみではなく、文の任意の長さのすべての部分に適用される採点方法を含む。グラフを使用するに当たっての重要な要素は、それが非常に多くの文に使用されていたとしても、サブツリーがたった一回だけ完全に採点され、分析されるということである。例えば、「Near the bank there is a bank」という句においては、「Near the bank」という句は、少なくとも二通りの意味を有するが、この句の最も適切な解釈はどれかという判断は一回

しか行われない。「there is a bank」という句も、同様に二通りの解釈を有するが、これら二通りの解釈のどれが最も適切であるかという判断は一回だけしか行われない。したがって、この文は異なる四つの意味に解釈できるが、サブ句のスコア付けは一回しか行われない。このグラフのもう一つの特徴は、各ノードが、文のその部分の長さに関して容易にアクセスできる情報で標識されていることである。この特徴により、英文を再度分析しなくても、その英文の任意のサブストリングのN個の最も適切な解釈を行うことができる。

しかし、一回の実行においては、その文の最も適切なN個の分析のみが、一回ごとに得られる（Nは20程度のある数）、グラフを使用することにより、システムは、もっと小さな構成部分についてのユーザの選択の結果を取り入れることができ、ユーザの選択を尊重したN個の異なる最も適切な分析を行う。これらすべての分析は、文の解析を再度行わず、または任意のサブストリングの採点を再度行わないで行われるので、迅速に行うことができる。

う構文特性を持つ。特性マッチングは、単語の特徴がそれらが適用される単語の特徴がある条件を満たした場合にだけ適用されるように、文法規則にマーク付けを行う。例えば、下記のような規則があり得る：

S = NP [⑥] + VP [⑥]

ここで、記号⑥はNPおよびVPの数の特性は一致しなければならないことを意味する。したがって、この規則は「they are」や「he is」は許容するが、「the y is」や「he are」は認めない。

特性マッチ制限は、「局所的」と「広範囲」とに分けられる。広範囲の動作は、実際に文が処理されるときにではなく、文法が作成されるときに計算され得る。次いで、実行しなければならない広範囲の一連の動作は、一連の命令バイトとしてコード化される。

広範囲特性動作の計算は、n成分の規則（すなわち、その右側に二つより多い入力を有し得る）でスタートしなければならない。その後、システムは、特性の集合が正しい様式で規則間を伝達して終わるように、種々の2成分の規則にコー

ドを割り当てる。n成分の規則を二つの要素からなる規則に分割することによって、解析操作は非常に単純化される。しかしシステムは、2成分の規則にまたがって特性の集合を追跡しているので、システムは広範囲操作のパワーを保持する。

本発明のシステムでは、複数の単語で構成する「熟語」も辞書の一部として許容し、一方、それらを構成している個々の（単語の）表示も保持する。これら二つの形態は、最終的には最も適切な表示であるために相互に競合し得る。例えば、「black sheep」の意味は、「持て余し者」の意味で辞書に見出される。しかし、ある場合には、「black sheep」という単語は、「黒い羊」を意味することもある。これらの形態は両方とも保持されているので、この熟語としてではない使用法も、なお正しい翻訳として選択され得る。

この熟語は、別の分類にも属し得る。例えば、システムは次の三つのタイプを使用し得る：

オールマイティ：United States of America

図8について説明すると、エキスパート評価部30の操作は、各翻訳を特徴づける種々の因子に基づいて行われ、これは種々のエキスパートにより処理される。確率規則エキスパート170は、最初のソース言語解析ツリーを入手するのに使用される文法規則の平均相対頻度を評価する。選択制限エキスパート178は、得られた翻訳の意味上の一致の度合を評価する。辞書見出し語確率エキスパート172は、最初のソース言語解析ツリーを入手するのに使用された文のいくつかの単語の特定の「品詞」の平均相対頻度を評価する。統計エキスパートは、ある翻訳のために選択した特定のフレーズの平均相対頻度を評価する。

システムは、種々の個別の英語の単語、句、および単語グループに対する英語の「品詞」（POS）を自動的に決定する。システムは、文を翻訳するとき、品詞を自動的に決定し、そしてシステムは通常は正しい決定を行う。しかし、ときには、翻訳する文それ自身が曖昧であることもある。1つより多い品詞として解釈し得る単語または句が含まれているときは、その単語又は句が出てくるその文について、複数の異なるが、しかしどれも「正しい」意味が得られることになる。システムのオペレータは、システムが自動的に決定する品詞を無視し、代わりに任意の単語や句または単語グループに対して手動で品詞を設定することも可能で

ある。例えば、「John saw a boy with a telescope.」というソース英文で、システムのオペレータが、「A boy with a telescope」を名詞句と見なすと、システムは、その文を「その少年は望遠鏡を持っていた」という意味に解釈し、従って「ジョンはその少年を見るために望遠鏡を使用した」という意味には解釈しないことになる。オペレータは、いくつかの可能な品詞設定を手動で行ったり、またはより制限のゆるやかな品詞設定を手動で行うことにより、システムが決定する品詞規則を上書きした場合、翻訳結果が悪化したり、少なくとも改善されないという状況を扱うこともあり得る。名詞句は名詞よりも制限がゆるやかであり、グループは最も制限が緩やかな品詞設定である。以下に種々の可能な品詞設定のリストを示す。

品詞 (POS)

名詞

名詞句  
動詞 (他動詞、自動詞)  
動詞句  
形容詞  
形容詞句  
副詞  
副詞句  
前置詞  
前置詞句  
接続詞  
グループ  
英語

「形容詞句」や「副詞句」の品詞設定は、ある英文が、システムが特定の前置詞句をどのように解釈するかによって意味が異なるような場合に有効である。例えば、「We need a book on the fourth of July」という文は、「On the fourth

of July」が形容詞としての意味を持つと解釈した場合には、「我々は米国の7月4日の独立記念日に祝する本がほしい。」という意味になるが、「On the fourth of July」を副詞句と解釈した場合には、「7月の4日に、我々は本がほしい。」という意味になる。システムが自動的に「On the fourth of July」に正しくない品詞を割り当てたとオペレータが考えた場合には、オペレータは、「We need a book on the fourth of July」という文の「On the fourth of July」に別の品詞を手動で設定し得る。オペレータが、システムに特定の単語、句または単語グループを英語から日本語に翻訳させたくないときは、オペレータは、所望の単語、句および/または単語グループに「英語」という品詞を設定し得る。オペレータは、設定がシステムにより自動的に行われた場合でも、オペレータにより手動で行われた場合でも、一つまたは複数の品詞設定を除去することもまた可能である。

システムは、ユーザごとに複数のレベルで訳語の使われ方の統計情報を追跡す

意図しない解析を区別するための効果的な方法でもある。

英語の多くの単語は、普通名詞としての解釈と固有名詞としての解釈とに潜在的な曖昧さを含んでいる。大文字エキスパート176は、大文字表記が意味を有する可能性がどの程度あるのかを決定するために、文中での大文字の位置を使用する。例えば、下記の文では、

Brown is my first choice.

My first choice is Brown.

第一の文は本来的に意味が曖昧であるが、第二の文は、「Brown」は色の名称ではなく、人名である可能性はるかに高い。このエキスパートは、大文字で始まる単語が文頭にあるのか、または文頭でない場所にあるのか(上の例)、その大文字で表示されている単語が辞書に含まれているかどうか、その小文字で始まる単語が辞書にあるかどうか、などの要因を考慮に入れる。このエキスパートは、文中の大文字で書かれた単語の正しい解釈を得るのに有効な方法である。

文が、最初が大文字の単語の連続を含んでいるとき、その文は、固有名詞または普通名詞の列として処理される。本発明のシステムは、大文字表記手順を使用しており、前者の解釈を優先する。上記列が、それ自身通常の文法規則により解析できないときは、その列は単一の解析していない名詞句として翻訳しないで処理される。この手順は、出現レベルの低い普通名詞の解釈を完全には無視しないで、複合の固有名詞を処理する非常に有効な手段であることが証明されている。

図7について説明すると、本発明の機械翻訳システムは、簡単な文法規則に基づく構造変換方法の効率を有しているが、テンプレート同構造変換方法の能力にも近い文法規則制御構造変換機構162を使用する。この方法は、平らでない複合構造を指定し得る文法規則160を使用する。以下は、他の翻訳システムで使用されている規則のフォーマットである；

$Y = > X1 + X2 + \dots + Xn$

指定のサブ構造

Y

る。例えば、システムは、単語の表面形式のレベル(「leaving」)が、自動詞としての使用に対して他動詞としての程度の頻度で使用されたか)で統計値を保持し、また意味レベル(「あとに残る」という意味か、または「～から出発した」という意味か)でも保持し、そして後者のタイプは「leave」、「leaves」、「left」、「leaving」の異なる変化形ごとに出現回数が積み上げられる。システムはまた、最後のいくつかの文で生じた使用についての統計値と、ユーザのサイトで任意の時間に生じた使用についての統計値とは区別して保持し得る。さらに、システムは、ユーザが単語の特定の意味を使用するように介入して指示した場合と、システムがユーザの確認なしに単語の特定の意味を使用した場合とを区別し得る。

構造バランスエキスパート182は、所定の文の構成部分の長さに関連する特徴で、英語や多くの他のヨーロッパ言語の特徴に基づいている。いくつかの(全てではないが)構造で、これらの言語では、軽い(短い)要素の左側に重い(長い)要素がくるような文は歓迎されない。例えば、

Mary hit Bill with a broom. (左が軽く、右が重い) [適合]

Mary hit with a broom Bill. (左が重く、右が軽い) [不適合]

Mary hit with a broom a dog that tried to bite her.

(左が重く、右がもっと重い) [適合]

所定の文の二つの解析がある場合、一方が「左が重く、右が軽い」列を避けようとする構造を含む「左が重く、右が軽い」列を含み、他の解析がそうでない場合には、前者はその文の意図する解析を要していないと見なされ得る。このエキスパートは、意図する解析と意図しない解析を区別するための有効な方法である。

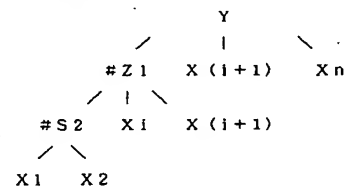
「A of B and C」というパターンの同格構造においては、意図する解析が「A of B and C」であるのか、または「A of B and C」であるのかを判断するのが難しい場合がある。同格構造エキスパート180は、どの同格モードが意味のより近い二つの要素を結合することになるかを決定するために、B C間の意味上の距離、およびA C間の意味上の距離を測定する。このエキスパートは処理中に意味特性ツリーにアクセスする。このエキスパートは、所定の文の意図する解析と

$X1 \quad X2 \dots Xn$

一方、本発明のシステムは以下のフォーマットの文法規則を使用する：

$Y = > \# Z1(i) \# Z2(2) X1 + X2 \dots + Xi + X(i+1) + \dots + X(n)$

指定のサブ構造



この構文で、「#」が前に付いている記号は、文の構造解析の目的では見えない仮想的な記号であるが、いったん所定の解析が入手できればサブ構造を構築するのに使用される仮想的な記号である。

このタイプの文法が与えられたとき、サブ構造において同格であるノードの任意の列の間で、複数の構造変換を特定することができるようになる。これにより、文法規則に基づく構造変換機構はテンプレート同構造変換機構のいくつかの能力を持つ機構に変換される。本発明のシステムは、上記の二番目のタイプの文法規則に基づいているけれども、一番目の形態の対応する文法規則を自動的にコンパイルする。したがって、文を解析するのに、第一の形態の文法規則を使用し、解析構造を形成するために第二の形態の文法規則を使用し得る。

構造変換は、また文法規則制御構造変換操作による操作をうけた後で、解析ツリー上で操作するために、辞書161にアクセスする辞書制御構造変換操作166を含む。次いで、生成規則制御構造変換操作が、標的言語テキスト41を供給するために、得られた解析ツリーに、生成規則を適用する。

再び、図1および図2について説明すると、システムが上記プロセスにより、最も望ましいとランク付けした翻訳を作成した後、その翻訳がディスプレイ20を通してユーザに提供される。次いで、ユーザは、その翻訳を採用するか、ユーザ入力装置22を通して別の解析システム37と相互作用させることによりその

翻訳を編集するかを選択する。編集操作中、ユーザは、正確に翻訳された翻訳文の部分は束縛しながら、その文の残りの部分の再翻訳を要求し得る。この操作は、迅速に行い得る。なぜなら、システムはエキスパートの重み31を含むグラフを保持しているからである。

図1-図9を参照して自然言語の自動翻訳システムの特定の詳細を説明してきたが、ここで、図10A-図16を参照して本発明による改善が記載される。これらの改善は、テキストがユーザ、オペレータまたはシステムによって発生されている間、たとえば、ユーザがワード処理のソフトウェアを使って文書を作成している間、その入力テキストを自動的に翻訳する。本発明のシステムの1つの利点は、入力テキストが発生されている間に、あるいはそのテキストが発生された直後に入力テキストを翻訳するためのシステムの能力にあり、この能力は言語学

習、および迅速且つ効率的に翻訳されなければならない一時的なテキストを発生するシステムにおいて有用であり得る。

本発明に従って、ユーザ、オペレータまたはシステムは、図1に示されているように、ユーザ入力インターフェース22、または入力インターフェース12によって、コンピュータ化されたシステムまたはファイルに、文字、記号およびテキスト（本明細書中以下で、文字、文字ストリーム、テキスト入力または入力テキスト・ストリームと交換可能に呼ばれる）を入力または追加することによって、テキスト・ファイルまたは文書を入力し、発生し、作成し、編集し、受信し、あるいはコンパイルする（本明細書中以下で、「テキスト・ファイル」を発生する）または「文書」を発生する」と交換可能に呼ばれる）ことができる。文書またはテキスト・ファイルは、所望の適用に依存して、任意の数の文字、フォーマット、テキストまたはファイルを含み得る。上記のように、ユーザ・インターフェース22および入力インターフェース12は、たとえば、キーボード、マウス、タッチスクリーン、ライトペン、音声活性化送信機または他の入力デバイス、ならびにモデムまたはシリアル・ポートなどのデジタル・インターフェースを含み得る。本発明のシステムは、ユーザのワークステーション（20、22）上にイン

入力テキストを翻訳および再翻訳するプロセス184が、ループとして示され、そのループは、受信されるたびに入力テキスト183を定期的に翻訳し、出力としてその翻訳されたテキストをディスプレイ185に対して送信し、そしてループ・バックして184において記憶部186から以前に受信された任意の入力された入力テキストと一緒に、新しく受信された入力テキストがあればそれを翻訳する。新しい入力テキストが受信されていない場合、翻訳プロセスは入力テキストがふたたび受信されるまで一時保留される。したがって、実質的にリアルタイムで、入力テキストが既存のテキスト・ファイルの中の入力テキストと一緒に翻訳され、ソース言語での入力テキストと、標的言語での出力テキストの両方を表示し得る。しかし、他の実施態様においては、翻訳されたテキストだけを表示する必要があり、ソース言語での入力テキストはその文書の中の翻訳された単語の上でクリックすることによって、あるいはシステムのクリップボードから得ることができることが理解される。

図10Bを参照すると、本発明の別の実施態様による自動翻訳および再翻訳の

ためのシステムを示している機能的ブロック図が示されている。上記と同様に、ユーザ入力インターフェース22および/または入力インターフェース12を通じてシステムの中に発生された、あるいは入力された入力テキスト183が翻訳エンジン16へ送られ、実質的にリアルタイムで記憶部18の中の入力テキスト・ファイルとして186において記憶される。次いで、その入力テキストが187においてその入力テキストが受信された時点までのテキスト・ファイルに対して比較され、その入力テキストのどの部分が以前に翻訳されていない新しく受信されたテキストであるかどうかを判定する。入力テキストが新しく受信された入力テキスト、たとえば、1つの文の継続、または前の文の修正を含んでいる単語の入力がどの部分であるかの判定が行われた後、その新しく受信された入力テキストについての翻訳188が実行される。新しく受信された入力テキストが現われる場所に依存して、たとえば、上記の例のような、新しく受信された入力テキストが以前に入力された文を修正している場合など、以前に受信された入力テキストの再翻訳188を実行し得る。このように、記憶されている入力テキスト186

ストールされている、例えば、MicrosoftのWord 6.0などのワード処理ソフトウェアを組み込むか、あるいはそれとインターフェースし得ることは理解されるべきである。したがって、ユーザがソース言語で文書が発生している時、標的言語へのその文書の自然言語翻訳が、たとえば、翻訳のアイコン上でクリックすることによってその入力テキストの翻訳を積極的にユーザが要求することなしに、自動的に実行される。しかし、本発明の翻訳システムは、積極的な翻訳要求を扱うこともできる。本発明のシステムは、翻訳されるべき文書がコンピュータ化されたシステムによって発生される場合に他のシステムとさらにインターフェースすることができる。文書は、ユーザの入力インターフェース22および入力インターフェース12を、独立に、あるいは組み合わせさせて発生させ得ることは理解されるべきである。

図1に関連して図10Aを参照すると、本発明の1つの実施態様による自動化された翻訳および再翻訳のためのシステムを示している機能的ブロック図が示さ

れている。示されているように、ユーザ入力インターフェース22および/または入力インターフェース12を通じてシステムの中に発生されたか、あるいは入力された入力テキスト183は、実質的にリアルタイムで翻訳エンジン16へ送られ、そこで184において、作成中のテキスト・ファイルまたは文書の中の入力テキストの始めから翻訳される。たとえば、さらに説明されるように、この翻訳は入力された文字のストリームの中の最初の文字、入力された単語および文字のストリームの中の最初の単語、または他の開始点から開始し得る。入力テキスト183はまた、それが発生されている時、あるいは入力されている時、その後の再翻訳184および表示185のために、記憶部18の中にテキスト・ファイルとして186の中にも記憶される。記憶された入力テキスト186をそのテキスト・ファイルの翻訳および再翻訳184において使用し得、その結果、即時受信された入力テキストおよび、既存のテキスト・ファイルの中に以前に入力された他の入力テキストが翻訳される。次いで、その翻訳および入力テキストが185において表示され、既存のテキストが新しく受信された入力テキストによって変化するたびに、絶えずリフレッシュされる。

を、テキスト・ファイルの翻訳および再翻訳184において使用して、即時受信された入力テキスト、および以前に入力された他の入力テキストが翻訳されるようにし得る。次いで、翻訳されたテキストおよび入力テキストの両方を185において表示し得る。入力テキストを翻訳および再翻訳するプロセス188がループとして示されており、このループは実質的にリアルタイムで新しく受信された入力テキストを翻訳し、その翻訳されたテキストを出力としてディスプレイに対して送信し、そして次に、ループ・バックして、その新しく受信された入力テキストを187においてテキスト・ファイルと比較し、新しく受信された入力テキストがあればそれを翻訳する。新しい入力テキストが受信されていなかった場合、翻訳プロセスを上記のように一時的に保留し得る。

ふたたび図1を参照すると、翻訳エンジン16の中に含まれている翻訳起動部21が、その文書の翻訳がいつ開始されるべきか、すなわち、翻訳エンジン14がソース言語から標的言語へのその文書の翻訳をいつ実行するかを決定する。図11を参照すると、本発明の1つの実施態様による翻訳起動部21がさらに詳細に示されている。示されているように、翻訳起動部21は処理モジュール200

を含み得、処理モジュール200は発振器と、タイマまたはカウンタ202（本明細書中以下で、「カウンタ202」と呼ばれる）とを含んでいて、カウンタ202はユーザ入力装置22または入力インターフェース12からの入力204において信号が存在していない時を検知する。たとえば、ユーザがタッチスクリーン190、キーボード192、またはマウス194を経由してタイプしている時、カウンタ202に対する入力204の1つまたはそれ以上がアクティブであり、カウンタ202はオフになっている。すなわち、カウンタ202はカウントしていない。入力204がアクティブでない時、あるいは入力204の特定のもののだけがアクティブである時、カウンタ202は開始値、たとえば、0からカウントを開始する。カウンタ202はそのカウント値を表している出力を比較部206に対して提供し、比較部206はそのカウント値を所定の値208に対して比較する。その所定の値は時間に関係しており、記憶部18の中に格納されている値であることが好ましい。例示の目的だけのために、本発明の翻訳システムが、

ユーザがそのテキストをワード処理文書の中にタイプしている時に翻訳するために使われている時、値208は約0.1秒から約4.5秒までの時間の期間を表し得、その期間は、たとえば、ユーザがどんなことを次にタイプしたいかに関して考えるため、あるいはその文書の中をスクロールするために一時休止する時間の長さにはほぼ等しい。さらに、本発明の翻訳システムが、コンピュータが発生したソースからの自然言語のテキストを翻訳するために使われる、本発明の別の実施態様においては、値208は上記より短いか、あるいはそれより長い時間の期間を表し得る。値208は、本発明のシステムの所望の使用に適している任意の時間に対応し得ることは理解されるべきである。

カウンタ202からのカウント値が、値208に実質的に等しいか、あるいはさもなければその値に対応している場合、比較部206は翻訳エンジン16に対して信号を提供し、翻訳プロセスを開始させる。そのカウント値が所定の値208に等しくないか、あるいはそれに対応しない場合、比較部206は加算部210に対して信号を提供し、カウンタ202をインクリメントさせる。従って、ユーザ入力装置22または入力インターフェース12からのテキスト入力の受信における遅延が存在した場合、カウンタ202はそのカウントが所定の値208に

実質的に等しくなるまでカウントし、その時点で翻訳が発生する。しかし、タッチスクリーン190、キーパッド192およびマウス194から翻訳エンジン16への入力196によって示されているように、エンター・キーの押下げ、あるいは翻訳アイコン上でのクリックなどのある種の動作によってもなお翻訳が発生させ得ることに注意することが重要である。

図12Aを参照すると、本発明の1つの実施態様に従って、文書が発生されている時にテキスト入力を翻訳するためのプロセスを示しているフローチャートが示されている。ステップ220に示されているように、翻訳起動部21はユーザ入力装置22または入力インターフェース12からの入力文字ストリームの受信における時間の経過を待つ。ある時間経過を検知すると、カウンタ202はステップ222に示されているように、初期値208からカウントを開始する。次に、ステップ224が実行され、次いで、翻訳起動部21は能動的に、あるいは受

動的に、即時受信された入力テキストを含んでいる更新されたテキスト・ファイルに対して227において比較され、即時受信したテキストを含んでいるテキストファイルについて判定が行われる。図10Bにおいて上で説明されたように、入力テキストはテキスト・ファイルとして記憶され、それが受信された時に翻訳される。入力テキストがその後受信された時、以前に記憶されている既存のテキスト・ファイルを調べ、それを即時受信された入力テキストと比較することによって判定が行われる。この方法で、その即時受信されたテキストだけ、あるいはその即時受信されたテキスト、および任意の他の必要なテキストが、次にステップ229において翻訳される。

図13Aおよび図13Bを参照すると、本発明の自然翻訳システムを使って、メモが発生される際に翻訳される例が示されている。図13Aに示されているように、ユーザがユーザ・インターフェースを通じて文字ストリーム250の中に、「The revised building permit」という単語をタイプした時、「permit」の後に一時休止が検知されている。その休止時間の間にユーザは文字を何も入力していない。これは普通はユーザが次に何をタイプしたいかについて考えている時に取る一時休止、あるいはコンピュータが発生した一時休止のためである。この時点で、翻訳起動部21は一時休止が所定の時間に等しいことを判定し、たとえば、3秒経過したことを判定し、英語の単語

の日本語への翻訳が、翻訳エンジンを使ってその文書の先頭から実行される。次に、その翻訳がディスプレイ20を経由してユーザに提示される。その翻訳が、プリンタを介して、ハード・コピーとして、あるいは音声シミュレーション送信機から発生される単語などの音声信号として、さらに提示され得ることは理解されるべきである。示されているように、このシステムは入力テキストが翻訳されたテキストと並べて提示され、ユーザが入力テキストを作成している間に出力のテキスト翻訳を見ることができるので、言語の学習のために非常に有用であり得る。

ここで図13Bを参照すると、ユーザがタイピングを再開した時、文字ストリ

動的に、キーボード・ストロークの存在を検知する。本明細書において、キーボード・ストロークはキーボード入力だけを指すことに限定されず、むしろ上記のコンポーネント（すなわち、キーボード、マウス、タッチスクリーン、ライトペン、音声活性化送信機）、および入力装置またはシグナリング装置として使われる他の装置のうちの1つまたはそれ以上を使っている、ユーザ装置22または入力インターフェース12からの任意の入力を含み得る。ステップ224からの出力が肯定的であった場合、ステップ220が実行され、そして翻訳起動部21は、ふたたび入力文字ストリームにおける時間経過を検知するために待つ。キーボード・ストロークが検知されない場合、ステップ226が実行され、翻訳起動部21はそのカウント値が所定の値208に達したかどうかを判定する。カウント値が所定の値208より小さかった場合、ステップ228が実行され、カウンタ202がインクリメントされ、そしてステップ224および226がふたたび繰り返される。

カウンタ202が所定の値208に達した場合、文書またはファイルにおける所定の場所、あるいは開始の場所、たとえば、文書またはファイルの先頭からソース文書の翻訳を開始し得る。本発明の実施態様においては、文書の先頭は、その文書が発生させる場合に、ユーザによって、あるいはコンピュータ・システム

によって入力または追加された最初の文字によって指定し得る。他の実施態様においては、所望の目的に依存して、最近入力された文、パラグラフまたはページの中に入力または追加された第1の文字または単語から翻訳を開始し得る。たとえば、ユーザが文書またはファイルの中に2ページのメモをタイプした場合、そのメモの1ページ目の最初の文字から翻訳を開始し得る。

図12Bを参照すると、本発明の別の実施態様に従って、文書が発生されている時にテキスト入力を翻訳するためのプロセスを示しているフローチャートが示されている。上記と同様に、翻訳起動部21は入力文字ストリームの受信における時間経過220を待ち、ある経過時間を検知すると、キーボード・ストローク224によって中断されない限り、カウント222を開始する。カウントが所定の値に達した時、以前に受信した入力テキストを含んでいる既存のテキスト・フ

ーム252が各キーボード・ストロークによって検知され、したがって、翻訳起動部21が待機モードへ戻る。次に、翻訳エンジンはユーザが「will be」をタイプした後の一時休止を検知し、最初に入力された文字、すなわち、「From: Richard Stevens...」から翻訳を開始させる。他の実施態様において、上記のように、文書の先頭ではなく、文の先頭から翻訳を開始し得ることは理解されるべきである。したがって、その翻訳は単語「The revised...」から開始されることになり、そして単語「permit」まで継続する。

文書の先頭から翻訳を開始することによって、すなわち、その文書を再翻訳することによって、中間の翻訳および最終の翻訳の精度を高め得ることに注目することは重要である。なぜなら、この翻訳エンジンは、その文の中の以前の単語の翻訳に影響を与える現在の文の構造におけるすべての変化、たとえば、その動詞に影響し得るそれ以降の変化を捕捉し得るからである。さらに、先行している単語、文またはパラグラフに対する変更、たとえば、切貼り機能を含む編集が行われている場合には、翻訳エンジンは標的の文書への翻訳がそのような編集を含むことを保証し得る。文の中のそれ以降の単語に対する変更が以前の単語の翻訳に影響する翻訳を例示するために、文の断片の翻訳と、その文の断片に対して単語が追加された後の文全体の翻訳とを比較しているフランス語の翻訳の以下の例を考える。

#### 入力I:

英語: I am

フランス語: Je suis

#### 入力II:

英語: I am not

フランス語: Je ne suis pas

したがって、翻訳が最初の文の終わった場所から開始された場合、すなわち、第2の入力の翻訳が「not」の後から開始された場合、フランス語の翻訳は正確でないことになる。なぜなら、それはフランス語における文の動詞の構造におい

て行われるべき変化を反映しないことになるからである。

図14を参照すると、テキスト入力を実際の翻訳するためのプロセスの別の実施態様が示されている。上記と同様に、ステップ320および322に示されているように、翻訳起動部21は時間経過を検知し、初期値208からのカウントを開始する。この実施態様においては、ステップ324において、キーボード・ストロークが検知された時、ステップ325が実行され、そしてそのキーボード・ストロークが、入力機能、マウスのクリック、タッチスクリーンの押下げ、または一時休止を表す他の入力信号または入力文字ストリームの入力に関係のない機能を表すかどうかについての判定が行われる。この実施態様においては、エンター・キーの押下げは、入力文字ストリームの発生における一時休止、たとえば、ユーザがパラグラフの終りに来て、最後のパラグラフと新しいパラグラフとの間にスペースを追加したい時の一時休止を表し得る。同様に、マウスのクリックは、たとえば、ユーザが次に続くテキストにアンダーラインを引くためにアイコン上でクリックした時、あるいはドラッグ・アンド・ドロップの機能を実行するためにテキスト上でクリックした時、入力文字ストリームの発生における一時休止を表し得る。そのような動作は入力テキストの翻訳が実行されるのに十分な時間を許すものとして翻訳起動部によって解釈され得る。すなわち、そのような例においては、翻訳起動部21は、そのキーボード・ストロークおよびマウス・クリックを一時休止と同様なものとし、ステップ330に示されているように、その文書の先頭からユーザが停止した場所の点まで、たとえば、その文書の中にタイプされた最後の単語または文字までの翻訳を実行させる。あるいは、上

記のように、即時受信されたテキストだけを翻訳し得る。

本発明の別の実施態様においては、翻訳起動部21は、図14のフローチャートが翻訳についての肯定的な要求を検知するように構成し得ることに注意することが重要である。すなわち、ステップ325において、システムは、ユーザ、オペレータまたはシステムが、エンター・キーの押下げ、アイコン上のクリック、タッチスクリーンの押下げまたは翻訳の要求を示す入力信号の発生によって文書の翻訳を要求しているかどうかを判定し得る。そのような実施態様において、

止または入力文字ストリームの入力に関連しない機能を表しているエンターまたはマウス・クリックなどの入力を受信されたかどうかを、ステップ425において必要に応じて判定し得る。上記と同様に、そのような入力を検知することによって、ステップ430が実行され、そこで文書の先頭、あるいは即時受信されたテキストの先頭の文字からなどの、その文書の中の開始点から翻訳が実行される。

他の実施態様においては、本発明の翻訳起動部21は、ステップ422において、ソース言語での認識可能な単語が発生されたかどうかを判定し得る。この実施態様においては、認識可能な単語、たとえば、英語の単語「av」ではない「an」が肯定的な判定を得る結果となり、ステップ424へ移る。「av」などの認識不可能な単語の場合、翻訳起動部21はステップ423において追加の文字を待ち、たとえば、ユーザがタイプし続けて「avenue」などの単語が発生された場合のように、発生されたかどうかを判定し得る。あるいは、翻訳起動部はステップ422において、たとえば、正しく入力された文字または単語がその後に続くタイプ誤りの場合のように、認識不可能な文字または単語の後に発生された文字または単語が認識可能であるかどうかを判定し得る。別の実施態様においては、翻訳起動部21はステップ424において、1つのパラグラフが入力文字ストリームによって形成されたかどうかを判定し得る。そのような実施態様においては、翻訳エンジン14は、入力文字ストリームがパラグラフを発生した後でのみ翻訳動作を実行する。

図13Cを参照すると、本発明の翻訳システムの一例が示されており、その中で翻訳起動部21がタイピングにおける時間経過の間に、完全な文がタイプされたかどうかを検知し、完全な文だけを翻訳するように構成されている。この実施態様の中で示されているように、パラグラフ254を含んでいる文は完全な文であり、そして翻訳されている。しかし、文の断片256「This should not delay the...」は、まだ翻訳されていない。なぜなら、これらの単語は単に文の断片を含むにすぎないことが翻訳エンジン16によって判定されているからである。

キーボード・ストロークまたは翻訳アイコン上のクリックの結果、その文書の中のテキストの自動翻訳および再翻訳が行われる。

ふたたびステップ324を参照すると、そこでは検知されたキーボード・ストロークが、エンター、マウス・クリックまたは上記のような他の入力以外のものであり、ステップ326が実行され、カウンタ202が所定の値208に達したかどうかに関する判定が行われ得、達していた場合、ステップ330が実行され、その文書の翻訳が開始点から発生される。カウンタ202が所定の値に達していなかった場合、ステップ328が実行され、カウンタがインクリメントされて、ステップ324および326がふたたび繰り返される。

図15を参照すると、本発明の別の実施態様が示されており、その中で翻訳起動部21が入力文字ストリームの翻訳を開始する前に、文の形態を検知する。上記と同様にステップ420において、翻訳エンジン16が入力文字ストリームを受信して、翻訳起動部21が一時休止を検知する。ステップ422において、翻訳起動部21は、その文字ストリームを調べて、文または文の断片（たとえば、1つの文字、文字の列、単語、または単語の列を含み得る）が、その入力文字ストリームによって形成されているかどうかを判定する。この実施態様においては、文または文の断片がまだ形成されていない場合、ステップ423において入力文字ストリームが待たれる。ふたたび入力文字ストリームが検知されると、ステップ420が実行され、そこで翻訳起動部21が、入力文字ストリームにおけるそれ以降の時間経過を待つ。文または文の断片が形成されたことが、ステップ422において判定された場合、ステップ424が実行され、その中で自動的に、あるいは受動的に、キーボード・ストロークが検知されたかどうかについての判

定が行われる。キーボード・ストロークが検知されなかった場合、ステップ430が実行され、その文書が開始点から翻訳される。キーボード・ストロークがステップ424において検知された場合、ステップ420が実行され、そこで翻訳起動部21がふたたび入力文字ストリームにおける時間経過を待つ。点線で示されているように、翻訳起動部21は、上記の図14で説明されたように、一時休

ここで図16を参照すると、本発明の1つの実施態様が示されており、それは、所定のカウンタ値280が到達されたかどうかを判定する概念、および文が入力文字ストリームによって形成されているかどうかを判定する概念を組み合わせている。上記の説明と同様に、翻訳起動部21は、ステップ520において、入力文字ストリームにおける時間経過を待ち、そしてステップ522において初期値280からのカウントを開始する。次に、翻訳起動部21はステップ524においてキーボード・ストロークが自動的に、あるいは能動的に検知されたかどうかを判定し、検知された場合、ステップ520がふたたび実行され、そして翻訳起動部21は入力文字ストリームにおける時間経過を待つ。キーボード・ストロークが検知されない場合、ステップ526において、カウンタ202が所定の値208に達したかどうかについての判定が行われる。カウンタ202が、そのような値280をまだカウントしていない場合、ステップ528がふたたび実行され、ステップ524および526がふたたび繰り返される。カウンタ202が所定の値208に達していたと判定された時、ステップ529が実行され、そこで翻訳起動部21は入力文字ストリームが少なくとも1つの文を形成するかどうかを判定する。入力文字ストリームが文を形成しない時、ステップ532が実行され、そこで入力テキスト・データの追加のストリームが待たれる。一旦、追加の入力文字ストリームが検知されると、制御がふたたびステップ522へ移り、そこで翻訳起動部21が入力文字ストリームにおけるそれ以降の時間経過を待ち、ステップ522から529までがふたたび実行される。ステップ529において、文字ストリームが文を形成するという判定がなされた場合、ステップ530が実行され、そしてその文書が開始点から翻訳される。本発明の他の実施態様においては、ステップ529は単語、文の断片、またはパラグラフをその入力テキストの形成を翻訳する前に検知し得ることは理解されるべきである。

本発明の自動翻訳および再翻訳システムは、標的言語を学習するためのソース言語において人を流暢にし得るための、言語学習ツールとして使用し得る。たと

えば、図13A～図13Cに示されているように、英語に堪能な人は、本発明の翻訳システムを容易に使用して日本語を学習し得る。上記で示されたように、デ

ディスプレイ・モニタを備えているコンピュータ・ワークステーションにいるユーザは、1つの単語または複数語の文をタイプして、ディスプレイ画面上でそのタイプされた単語の自動翻訳を見ることができる。さらに、ユーザがタイプし続けると、そのユーザは文法および他の構造的な影響のために先行している単語の翻訳において発生し得る変化を見ることができる。

本発明のシステムは、ユーザが一掃に見るために、1つのソース言語から複数の標的言語への翻訳、たとえば、英語から日本語へ、そして英語から韓国語へ、あるいは英語から日本語へ、そして日本語から韓国語への翻訳を扱うように構成でき、ことに注意することがさらに重要である。

さらに、本発明のシステムは、例えば、多数の異なる国にいる多数のネットワーク・ユーザに対して知らせる必要がある重要な、あるいは緊急のメッセージを、その関連、適用性、またはフォーマットにおいて一時的であるコンピュータ発生の自然言語テキストを翻訳するために使用し得る。そのような例において、本発明のシステムは、そのメッセージが発生されている時に1つまたはそれ以上の言語にそのメッセージを翻訳し、そのメッセージが関連のユーザに対して迅速に到達し、理解されるようにするために使うことができる。

本発明は、さらにインターネット関連のアプリケーション、たとえば、例示の目的のためにいくつか例を挙げれば、インスタント・メッセージ、電子メール、ウェブ・ページ、掲示板およびインターネットのチャット・ルームなどの適切なアプリケーションにおいて有効に使うことができる。しかし、その他の多数のアプリケーションが本発明の適用範囲内にあることが理解されるべきである。

上記の機能および操作のすべてを、各種の固定配線の回路設計によって、および/または汎用のコンピュータで使うためのプログラミング技法によって実装し得る。フローチャートに示されているようなステップは、そこに示されている順序で一般的に適用される必要はなく、ステップの組合せを組み合わせて得る。同様に、このシステムの機能を、プログラムおよびデータ中に各種の方法で分配し得る。さらに、文法および他の動作規則を1つまたはそれ以上の高級言語で開発し

一方、それらをコンパイルされたフォーマットでエンド・ユーザに対して供給するのが有利な場合がある。

本明細書中に記載された機能のすべてを含んでいる、本明細書中に記載された自動自然言語翻訳システムの任意の実施形態を、汎用のコンピュータ（たとえば、AppleのMacintosh、IBM PCまたはその互換機、Sunのワークステーションなど）で実行させるためのディスクまたは光コンパクト・ディスク（CD）などの、コンピュータ読み取り可能媒体上のコンピュータ・ソフトウェアとして提供し得る。

この分野の当業者であれば、請求の範囲に記載されている本発明の精神および範囲から逸脱することなしに、本明細書中に記載されることの変更、修正、およびその他の実施を思い付くことができる。したがって、本発明は、前記説明的な記述によってではなく、以下の請求項の精神および範囲によって規定されるべきである。

【図1】

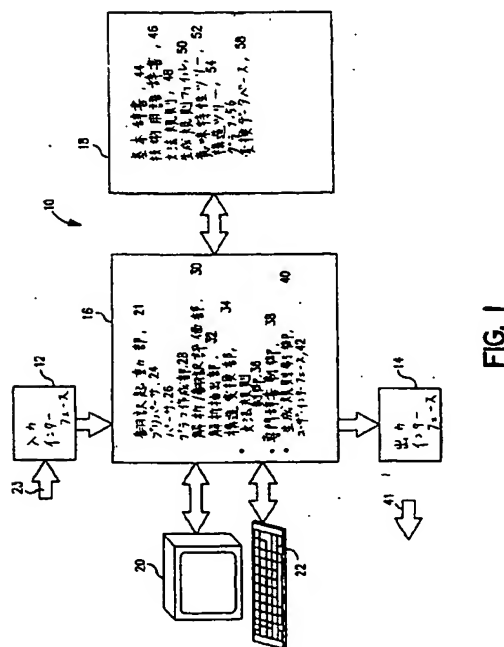


FIG. 1

【図2】

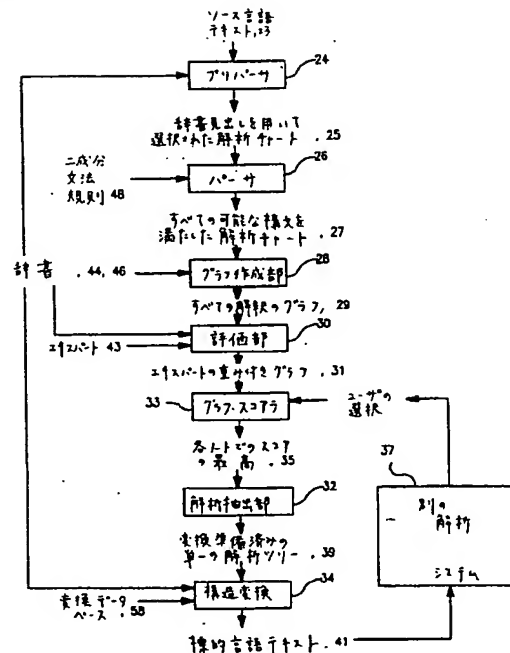


FIG. 2

【図3】

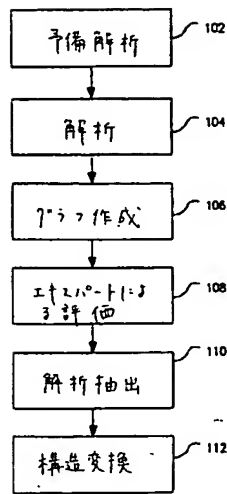


FIG. 3

【図4】

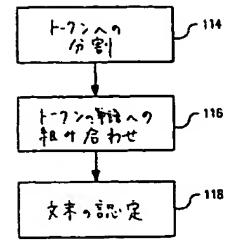


FIG. 4

【図5】

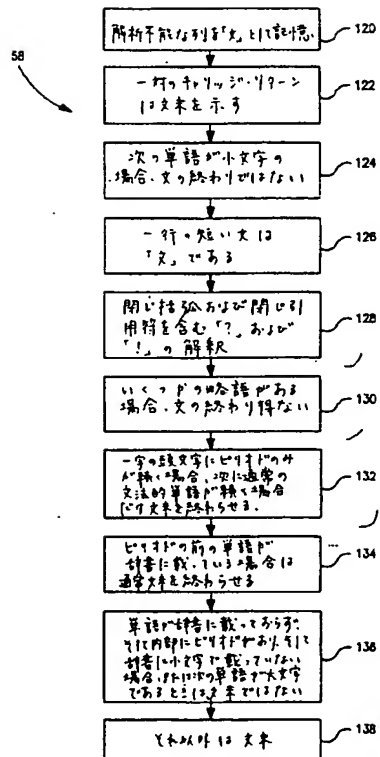


FIG. 5

【図6】

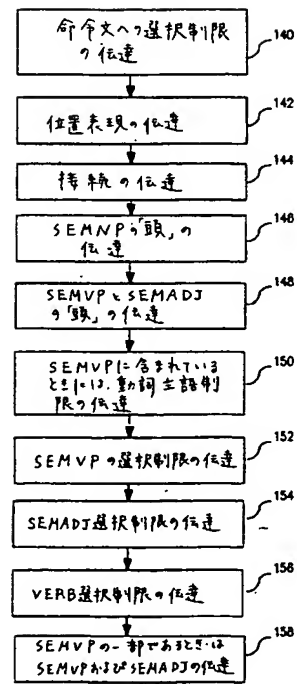


FIG. 6



【図7】

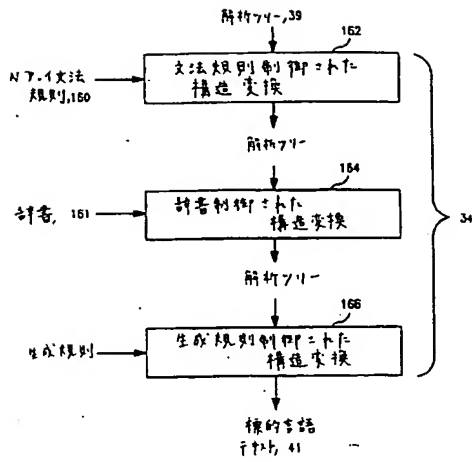


FIG. 7

【図8】

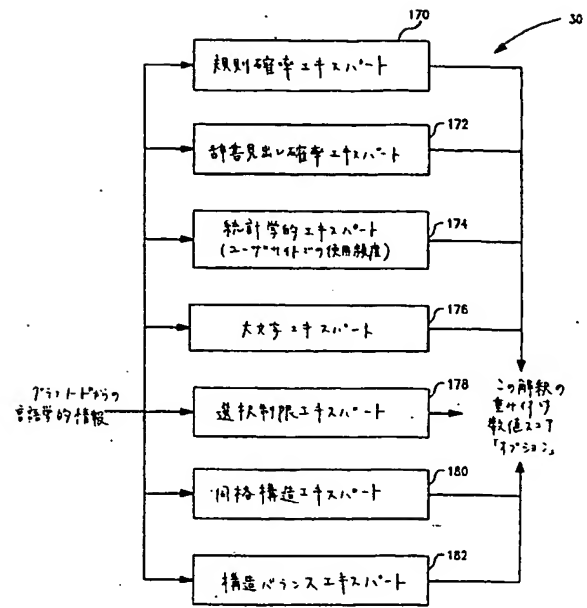


FIG. 8

【図9】

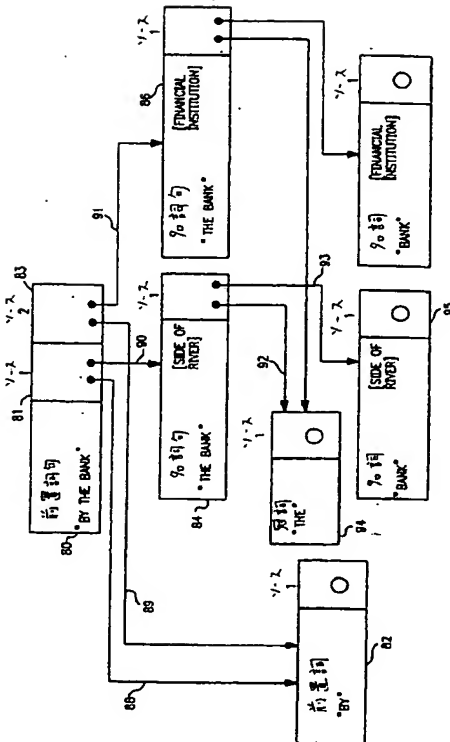


FIG. 9

【図10】

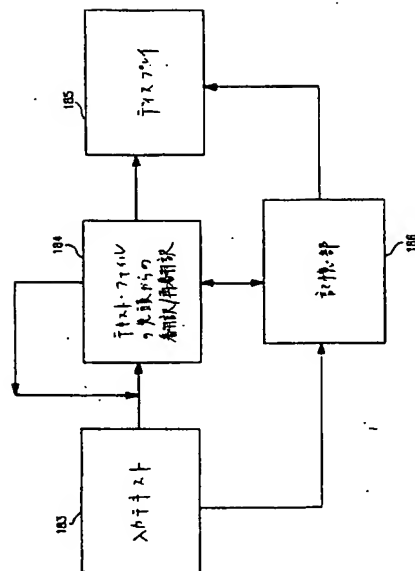


FIG. 10A

【図10】

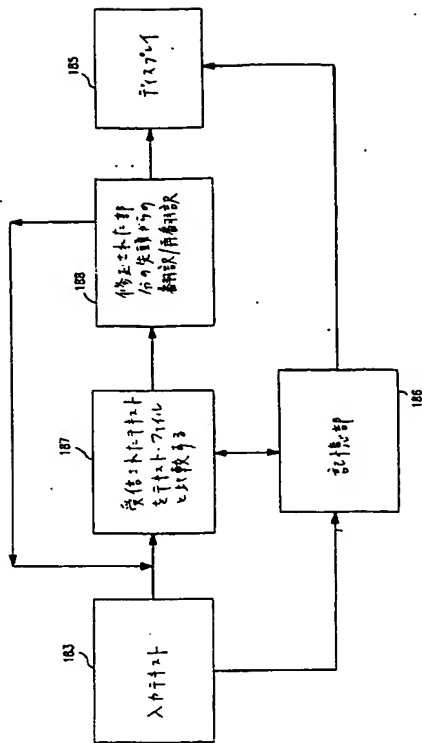


FIG. 10B

【図11】

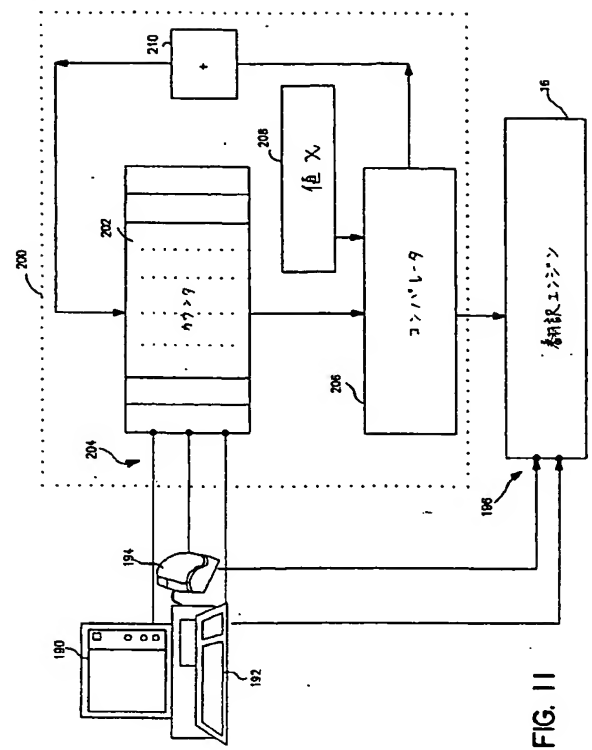


FIG. 11

【図12】

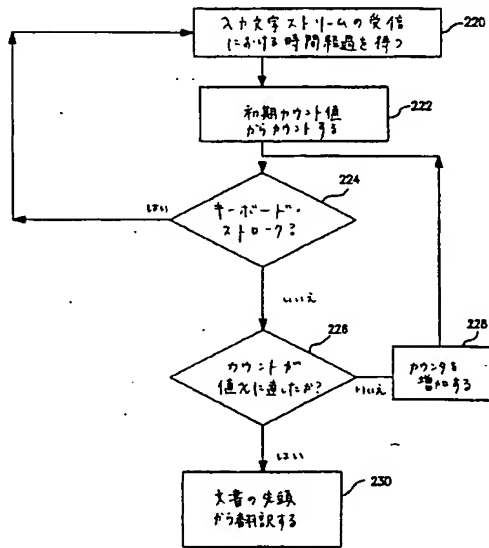


FIG. 12A

【図12】

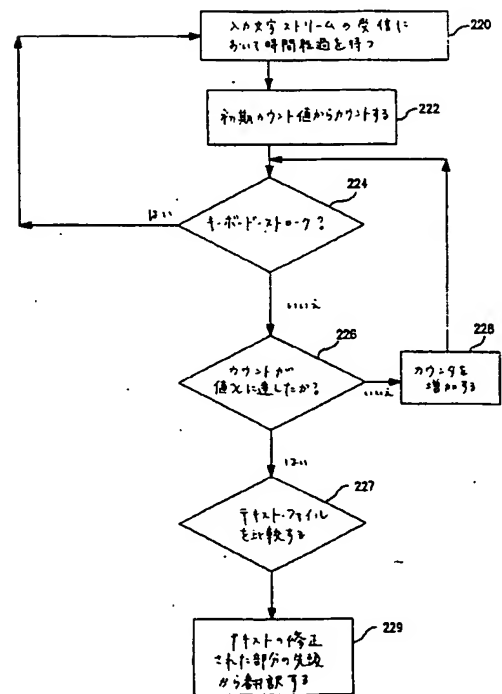


FIG. 12B

図13

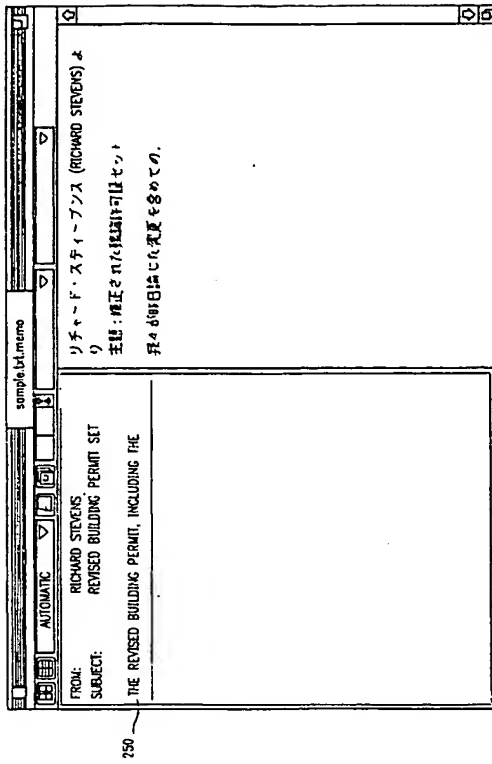


FIG. 13A

図13

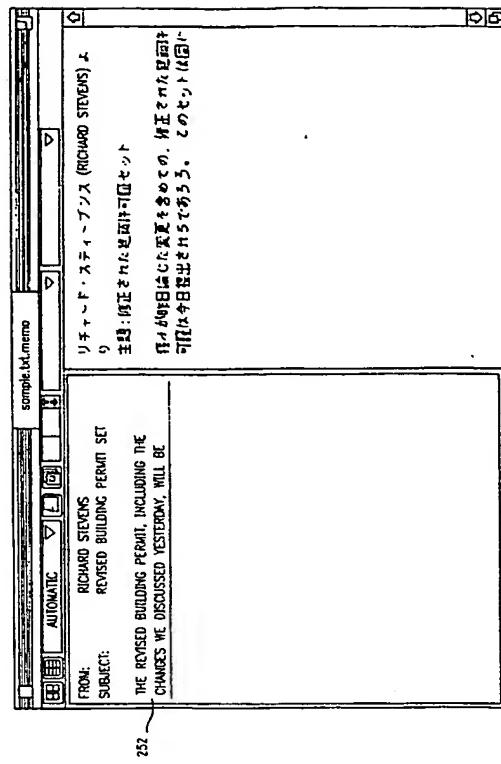


FIG. 13B

図13

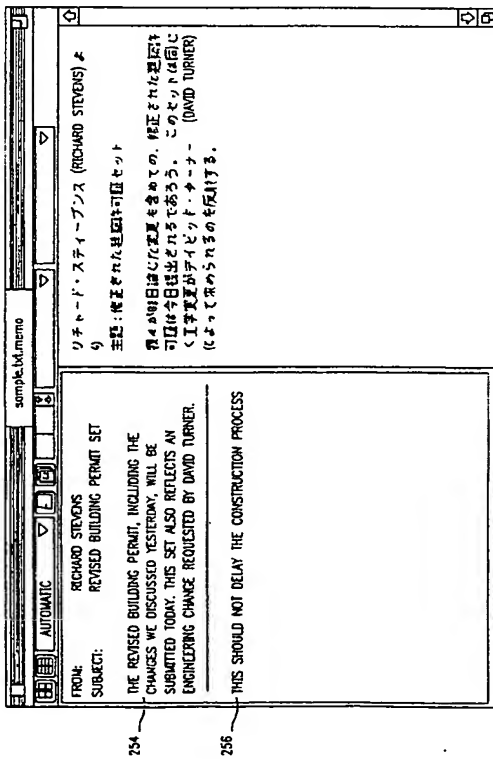


FIG. 13C

図14

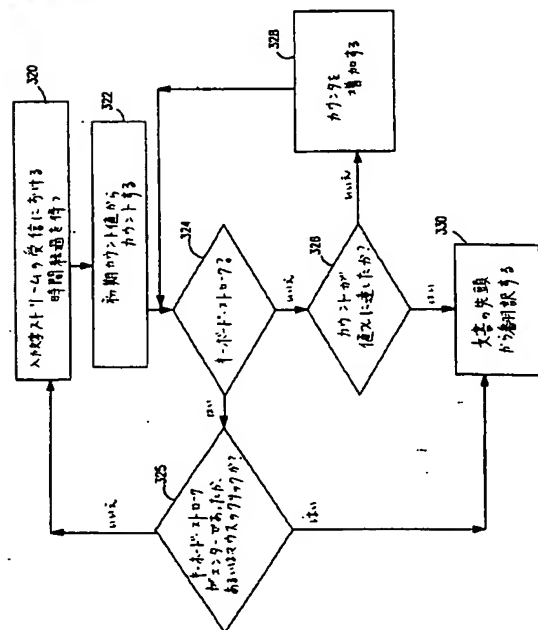


FIG. 14

【図15】

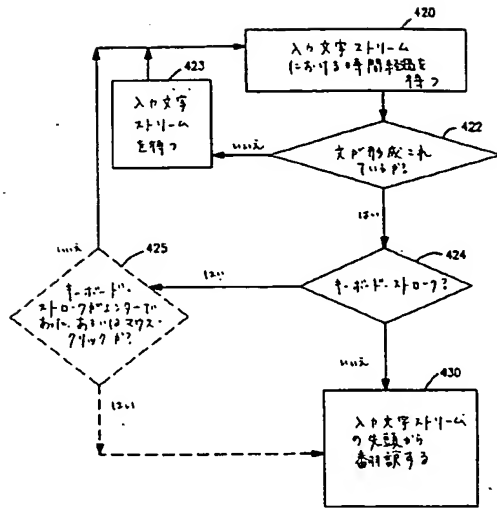


FIG. 15

【図16】

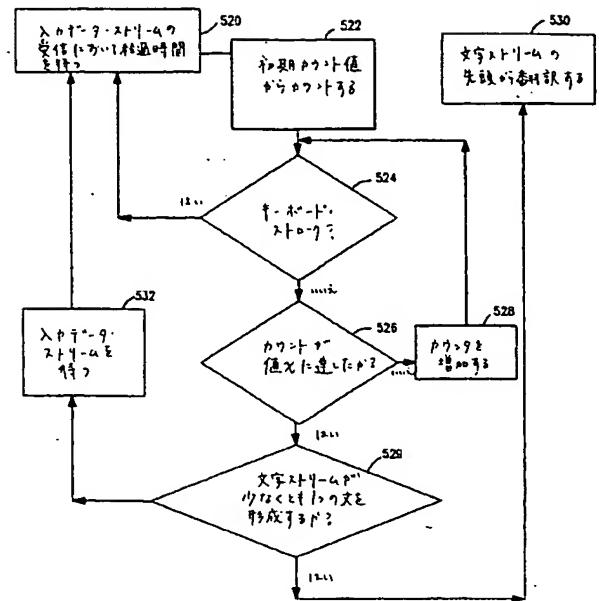


FIG. 16

## 【手続補正書】

【提出日】平成12年2月21日(2000. 2. 21)

## 【補正内容】

## 請求の範囲

1. 自動自然言語翻訳システムであって、以下：

メモリ・モジュール、

該メモリ・モジュールと通信する受信モジュールであって、ソース自然言語での入力テキスト情報を受信し、そして該入力テキスト情報を該メモリ・モジュールの中に格納する、受信モジュール、

該受信モジュールと通信する翻訳エンジンであって、該メモリ・モジュールにアクセスし、そして該ソース自然言語での該入力テキスト情報を標的自然言語での出力テキスト情報に翻訳する、翻訳エンジン、ならびに、

該受信モジュールと通信する処理モジュールであって、該受信モジュールが入力テキスト情報を受信していないことを該処理モジュールが決定するときに、該格納されている入力テキスト情報を翻訳するように該翻訳エンジンに対して命令信号を自動的に送信する、処理モジュール、を含むシステム。

2. 請求項1に記載のシステムであって、前記ソース自然言語が英語であり、そして該標的自然言語が日本語である、システム。

3. 請求項1に記載のシステムであって、前記受信モジュールと通信し、前記ソース自然言語での入力テキスト情報を提供するユーザデバイスをさらに含む、システム。

4. 請求項3に記載のシステムであって、ユーザが前記ユーザデバイスを使って前記ソース自然言語での文書を作成しているときに、前記受信モジュールが該ユーザデバイスから入力テキスト情報を受信するシステム。

5. 請求項3に記載のシステムであって、前記ユーザデバイスが、コンピュータ・キーボード、マウス、タッチスクリーン、または音声活性化送信機の中の

少なくとも1つを含む、システム。

6. 請求項1に記載のシステムであって、該受信モジュールが文字でない入力を表す信号を受信したことを判定することによって、前記処理モジュールが前記受信モジュールが入力テキスト情報を受信していないことを判定する、システム。

7. 請求項1に記載のシステムであって、該受信モジュールがマウスのクリックを表す信号を受信したことを判定することによって、前記処理モジュールが前記受信モジュールが入力テキスト情報を受信していないことを判定する、システム。

8. 請求項1に記載のシステムであって、該受信モジュールが所定の期間入力テキスト情報を受信していないことを判定することによって、前記処理モジュールが前記受信モジュールが入力テキスト情報を受信していないことを判定する、システム。

9. 請求項1に記載のシステムであって、前記翻訳エンジンと通信して、前記ソース自然言語での前記入力テキスト情報および前記標的自然言語での前記出力テキスト情報を表示するディスプレイをさらに含む、システム。

10. 請求項8に記載のシステムであって、前記処理モジュールが、カウント値が前記所定の期間に等しくなるまで、該カウント値を増加するためのカウンタを、さらに含む、システム。

11. 請求項10に記載のシステムであって、前記受信モジュールが入力テキスト情報を受信するときに、前記カウンタがリセットされる、システム。

12. 請求項1に記載のシステムであって、前記翻訳エンジンが、前記指示信号の送信時に先立って、前記受信モジュールによって受信された入力テキスト情報を翻訳する、システム。

13. 請求項1に記載のシステムであって、前記入力テキスト情報が入力デバイスから送信される文字のストリームを含む、システム。

14. 請求項8に記載のシステムであって、前記翻訳エンジンに対して前記指示信号を送信する前に、前記所定の期間に等しい、前記入力テキスト情報における時間経過の存在を判定することによって、前記処理モジュールが前記受信モジュールがテキスト情報を受信していない場合を判定する、システム。

15. 文字ストリームの自然言語翻訳を自動的に実行する方法であって、以下のステップ:

複数の文字を含んでいるソース言語での第一の文字ストリームをユーザデバイスから受信するステップ、

該第一の文字ストリームの受信において時間経過が存在するか否かを判定するステップであって、該時間経過の間に文字は全く受信されない、ステップ、および

該時間経過が存在する場合、該ソース言語から標的言語へ該第一の文字ストリームの部分を翻訳するステップ、  
を包含する、方法。

16. 請求項15に記載の方法であって、以下のステップ:

前記時間経過が所定の期間に等しいかどうかを判定するステップ、および

該時間経過が該所定の期間に等しい場合、前記第一の文字ストリームの中の最初の文字から始めて、前記ソース言語から前記標的言語へ該第一の文字ストリー  
ムを翻訳するステップ、

をさらに包含する、方法。

17. 請求項16に記載の方法であって、前記所定の期間が約0.3秒〜2秒の範囲内にある、方法。

18. 請求項15に記載の方法であって、前記第一の文字ストリームが前記ソース言語での少なくとも1つの単語を含む、方法。

19. 請求項15に記載の方法であって、前記第一の文字ストリームが前記ソース言語での少なくとも1つの文を含む、方法。

20. 請求項15に記載の方法であって、以下のステップ:

コンピュータ・キーボードを使って前記第一の文字ストリームを発生するステップ、および

前記ソース言語での該第一の文字ストリーム、および前記標的言語での前記翻訳された文字ストリームをコンピュータ・モニタ上で表示するステップ、  
をさらに包含する、方法。

新たに受信したテキストを含む該第二の文字ストリームの部分を前記標的言語に翻訳するステップ、

をさらに包含する、方法。

26. 請求項24に記載の方法であって、前記第二の文字ストリームを、前記標的言語に翻訳するステップをさらに包含する、方法。

27. 請求項20に記載の方法であって、前記文字ストリームを表示するステップが、前記標的言語における文字の各々が生成されるときに前記翻訳された文字ストリームを表示することを包含する、方法。

28. 請求項8に記載の方法であって、前記所定の期間が、0.3秒〜1秒の範囲内である、方法。

21. 自動自然言語翻訳システムであって、以下:

受信モジュールであって、ユーザワークステーションとインターフェースし、かつ該ユーザワークステーションにおいてユーザによって発生されたソース言語での文字ストリームを受信する、受信モジュール、

該受信モジュールと通信する処理モジュールであって、該文字ストリームにおける文字の受信において時間経過が存在するかどうかを判定する、処理モジュール、および

該処理モジュールと通信する翻訳モジュールであって、該時間経過が存在するときに該ソース言語での該文字ストリームを該標的言語での文字ストリームへ自動的に翻訳し、そして該翻訳された文字ストリームを該ユーザに送信する、翻訳モジュール、

を含む、システム。

22. 請求項1に記載のシステムであって、前記受信モジュールが、以前に前記標的言語に翻訳された入力テキストを受信する、システム。

23. 請求項1に記載のシステムであって、前記受信モジュールが以前に受信した入力テキストに対して修正を生じさせる入力テキストを受信する、システム。

24. 請求項15に記載の方法であって、以下のステップ:

メモリ・デバイス上に前記ソース言語における前記第一の文字ストリームを格納するステップ、

前記ユーザデバイスから、該ソース言語における複数の文字を含む第二の文字ストリームを受信するステップ、および

該第二の文字ストリームが、該第一の文字ストリームの少なくとも一部を改変する場合に、該第一の文字ストリームの少なくとも一部を再翻訳するステップ、  
をさらに包含する、方法。

25. 請求項24に記載の方法であって、以下のステップ:

前記第一の文字ストリームと、前記第二の文字ストリームとを比較して、該第二の文字ストリームの少なくとも一部が新たに受信したテキストを含むかどうかを判定するステップ、および

## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 G06F17/28		International Application No. PCT/US 98/00729
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 725 353 A (OKI ELECTRIC IND CO LTD) 7 August 1996 see abstract; claims 1-5 see page 8, column 20 - column 55; figure 4 ---	1, 15, 21
A	US 4 882 681 A (BROTZ GREGORY R) 21 November 1989 see abstract see column 4, line 41 - line 58 ---	1, 15, 21
A	"SPEECH-TO-SPEECH TRANSLATOR FOR NATURAL LANGUAGE" IBM TECHNICAL DISCLOSURE BULLETIN, vol. 37, no. 28, 1 February 1994, pages 137-140, XP000433789 see page 1 - page 3 ---	1, 15, 21
-/--		
<input checked="" type="checkbox"/> Further documents are cited in the continuation of box C.		<input checked="" type="checkbox"/> Patent family members are listed in annex
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "Z" document member of the same patent family		
Date of the actual completion of the international search 29 April 1998		Date of mailing of the international search report 08/05/1998
Name and mailing address of the ISA European Patent Office, P.B. 5616 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016		Authorized officer Suendermann, R

Form PCT/ISA/210 (second sheet) (July 1992)

## INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/US 98/00729

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 690 436 A (SEL ALCATEL AG ;ALCATEL NV (NL)) 3 January 1996 see abstract; claim 1 -----	1,15,21

Form PCT/ISA210 (continuation of second sheet) (July 1992)



## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 98/00729

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0725353 A	07-08-96	JP 8202734 A	09-08-96
		JP 8202721 A	09-08-96
		JP 8202722 A	09-08-96
		CA 2168395 A	01-08-96
US 4882681 A	21-11-89	NONE	
EP 0690436 A	03-01-96	DE 4422545 A	04-01-96
		AU 2328495 A	11-01-96
		FI 953176 A	29-12-95
		JP 8044386 A	16-02-96

Form PCT/ISA/210 (patent family annex) (July 1992)